MANGALORE UNIVERSITY

Final Synopsis

of the PhD Thesis to be submitted for the award of degree of
Doctor of Philosophy in

Computer Science

TITLE OF THE WORK

Machine Learning Algorithms for Pulmonary Disease Classification

By
Shazia Mannan
Research Scholar

Under the Guidance of
Dr. B. H. Shekar,
Senior Professor

Department of Post-Graduate Studies and Research in Computer Science,
Mangalore University,
Mangalagangotri-574 199

June 2025

Contents

1 Introduction **1**
- 1.1 **A Brief Introduction to Pulmonary Diseases** . . . 2
  - 1.1.1 **Tuberculosis** . . . 2
  - 1.1.2 **COVID-19** . . . 3
  - 1.1.3 **Viral Pneumonia** . . . 3
  - 1.1.4 **Lung Cancer** . . . 4
- 1.2 **Imaging Modalities and Dataset Description** . . . 5
- 1.3 **Literature Review** . . . 6
- 1.4 **Challenges in Pulmonary Disease Classification** . . . 10

2 Objectives **11**

3 Proposed Method 1: Hand-crafted Textural and Deep Features for Multi-Class Lung Disease Classification **11**
- 3.1 Methodology . . . 11
- 3.2 Results . . . 12
- 3.3 Discussion . . . 13

4 Proposed Method 2: Differential Diagnosis of Pulmonary Diseases using Convolutional Neural Network and its variants **15**
- 4.1 Convolutional Neural Network with LSTM Architecture . . . 15
  - 4.1.1 Methodology . . . 15
  - 4.1.2 Results . . . 16
  - 4.1.3 Discussion . . . 17
- 4.2 CNN-GRU Model for Tuberculosis Detection Enhanced with LIME Interpretability . . . 18
  - 4.2.1 Methodology . . . 18
  - 4.2.2 Results . . . 18
  - 4.2.3 Discussion . . . 20
- 4.3 Customized CNN and Oversampling Techniques: A Deep Learning Approach 21
  - 4.3.1 Methodology . . . 21
  - 4.3.2 Results . . . 21
  - 4.3.3 Discussion . . . 24

5 Proposed Method 3: Deep Ensemble Model for Differential Diagnosis of COVID-19 Pneumonia **25**
- 5.1 Methodology . . . 25
- 5.2 Results . . . 26
- 5.3 Discussion . . . 27

6 Proposed Method 4: An Integrated and Multitask Framework for Enhanced Segmentation and Classification of Pulmonary Diseases **27**
- 6.1 Integrated Segmentation and Classification Models for Pneumonia Detection 27
  - 6.1.1 Methodology . . . 27
  - 6.1.2 Results . . . 28
  - 6.1.3 Discussion . . . 29

6.2 PulmoMTL-Net: A Multitask Learning Framework for Enhanced Segmentation and Classification of Pulmonary Diseases 30
6.2.1 Methodology 30
6.2.2 Results 32
6.2.3 Discussion 32

7 Proposed Method 5: Performance Analysis of Deep Neural Networks and Machine Learning Algorithms for Pulmonary Disease Classification 34
7.1 Methodology 34
7.2 Results 36
7.3 Discussion 36

8 Conclusions and Future Works 38
8.1 Conclusions 38
8.2 Future Works 41

9 Publications 42

Machine Learning Algorithms for Pulmonary Disease Classification

1 Introduction

Pulmonary diseases, also known as lung or respiratory diseases, are illnesses that affect the lungs' air shaft and other associated organs. Viral Pneumonia (VP), Tuberculosis (TB), Lung Cancer, Asthma and Corona Virus Disease (COVID-19) are some examples of pulmonary diseases. The first three diseases are responsible for moderate to serious illness and mortality in people with lung infection around the world[23]. Covid-19 has been declared as a pandemic and because of its highly contagious nature, is spreading unhindered in the world. According to Forum of International Respiratory Societies [26], about 334 million people suffer from asthma, and, each year, tuberculosis kills 1.4 million people, 1.6 million people die from lung cancer, while pneumonia also kills millions of people. It is clear that lung diseases are one of the leading causes of death and disability in this world. Early detection plays a key role in increasing the chances of recovery and improve long-term survival rates. Symptoms for VP, TB and COVID-19 are very similar in nature. They include: fever, chills, cold and cough, dyspnea (shortness of breath), chest pain that happens when a person breathes deeply or coughs, fatigue, diarrhea, myalgia (body pain) and so on. TB and COVID-19 are both infectious disease that primarily attacks the lungs and their main transmission route is via respiratory droplets[9],[33]. Similarly, Lung cancer manifests symptoms such as chest pain, shortness of breath, fatigue, myalgia, hoarseness and others. Radiological manifestations such as ground-glass opacities (GGOs) and crazy paving patterns are found in diseases such as COVID-19 especially in the beginning of the ailment. This could lead to confusion in differential diagnosis with other pathologies such as the initial stage of lung cancer.

Traditionally, pathological tests such as blood tests and culture, sputum test, biopsy, skin tests such as Mantoux and radiological examinations are carried out to detect pulmonary diseases. Manual diagnosis of diseases is expensive, prone to human errors and time consuming[12]. In case of radiological examinations such as CXR and CT-Scan, medical professionals interpret the biomedical images according to their individual judgement. The correct analysis and interpretation of the medical images is the key to successful diagnosis and subsequent therapy. Till date doctors have been relying on their past experiences to extract and classify the images into appropriate classes or groups. This leads to unnecessary delays and difficulties [39],[15]. It is further seen that doctors are restricted by their speciality and cannot correctly analyze images related to other fields. Artificial Intelligence techniques like Machine Learning (ML) and Deep Learning (DL) can instinctively recognize patterns in the biomedical image data and have produced excellent results[40]. With advancements in technology, numerous medical imaging such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), positron imaging tomography (PET), ultrasound imaging etc., can be easily acquired for training a

computer-aided model to detect and identify various pathologies with speed and accuracy. Computer-aided diagnostic(CAD) systems provide a great help to the medical fraternity by categorizing biomedical images to their relevant classes, thereby reducing the workload for health professionals. Recent developments in ML, particularly Deep Learning, support the identification, quantification and classification of patterns in medical images. These developments were made possible due to the ability of DL to learn features merely from data, instead of hand-designed features based on domain-specific knowledge. DL is quickly becoming state of the art, leading to improved performance in numerous medical applications. Consequently, these advancements assist clinicians in detecting and classifying certain medical conditions efficiently [40].

Numerous datasets (CXR and CT images) have been released for public use so that researchers can utilize them in their algorithms and research studies. Chest radiographs depicting COVID-19, CAP and various other pulmonary diseases can be obtained from these databases. Researchers employ datasets comprising of data for numerous diseases to show for the models' generalization ability. By working with such diverse image data it could be proved that the model is capable of accurately identifying a specific ailment. Some of these datasets contain CXR images and CT-scan images of COVID-19, while others include those of normal subjects and different pulmonary diseases like TB, VP or Lung cancer. The reason for using the latter type of datasets is to create more generalizable algorithms that can detect COVID-19 from a pool of more diverse radiography images[2].

The area of lung/pulmonary disease detection is plagued by a number of challenges, as identified from the previous research studies that were considered. There were five major concerns identified: i) Data imbalance; (ii) lack of annotated and labeled images; (iii) Limited available datasets; and (iv) When using Ensemble techniques, high correlation of errors;(v) Handling of large image size; (vi)noisy images; (vii) In most instances, chest/lung image classification and segmentation are thought to be a binary process, with present vs absent, standard versus abnormal being the most common examples. In many cases, for example, a patient with one chest ailment may have multiple chest diseases. As a result, the model often performs less properly. The majority of the researchers chose to avoid this ambiguity because training with such ambiguous data is difficult. DL approaches for multi-class classification problems are a common answer to this problem[23],[17].

1.1 A Brief Introduction to Pulmonary Diseases

1.1.1 Tuberculosis

TB, a disease that dates back thousands of years, is caused by Mycobacterium tuberculosis, which was discovered in 1882 by Dr. Robert Koch [36]. While TB is both treatable and curable, it remains a deadly affliction that causes fatalities in millions[22]. In 2022, TB claimed the lives of 1.3 million individuals worldwide, positioning it as the second most significant infectious cause of death following COVID-19 [31]. The emergence of COVID-19 in 2019 further exacerbated the condition of suspected/confirmed TB patients due to enforced lockdowns, restricted access to hospitals, and difficulties in obtaining life-saving medications. Without treatment, TB has a high mortality rate, emphasizing the critical need for early detection and medical assistance. According to the statistics provided by India TB Report 20241, compiled by the health ministry, 2.55 million Indians have been diagnosed with TB in 2023 and the mortality rate is 0.32 million. Sputum smear microscopy, though time-consuming remains the primary diagnostic tool for TB globally due to its affordability and simplicity.

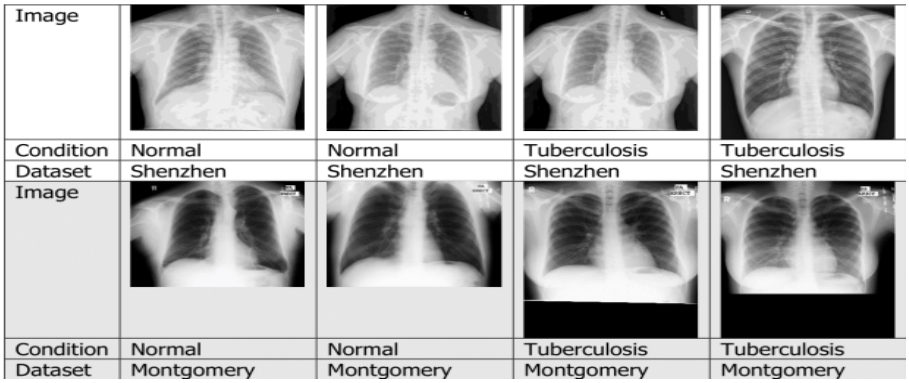

Figure 1: Sample images from Tuberculosis CXR datasets.

1.1.2 COVID-19

COVID-19 pneumonia, also known as novel coronavirus-infected pneumonia (NCIP) is a clinical manifestation of SARS Coronavirus that leads to inflammation and clogging up of the lungs' alveoli with debris and mucus. Since its outbreak in December 2019 till 06 November 2024, there have been 704,753,890 confirmed cases of COVID-19, including 7,010,681 deaths around the world. The diagnostic indications of COVID-19 include fever, cough, dyspnea, anosmia, myalgia, etc. The clinical tools being employed presently include RTPCR and antigen testing kits, which work by detecting the Sars-Cov2 virus from specimens taken from nasopharyngeal or throat swabs[45].

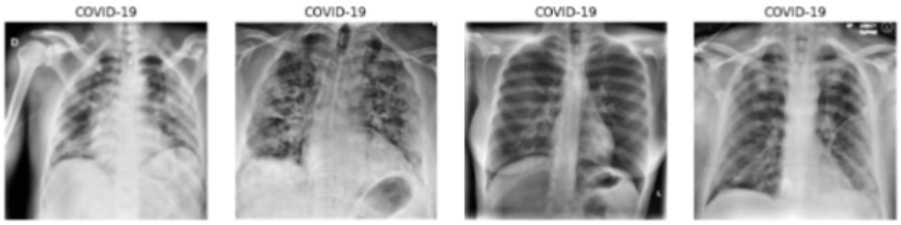

Figure 2: Sample COVID-19 images from C19RD dataset.

1.1.3 Viral Pneumonia

Pneumonia is the primary cause of loss of life in paediatric cases worldwide. According to the World Health Organization (WHO), this infectious disease proved fatal for 740180 pre-schoolers in 2019 along with a large number of adults. It accounted for 14% of all deaths of toddlers and preschool children. During pneumonia, air cells that make up the lungs become saturated with pus and fluid, causing breathing to become laboured and difficult, resulting in a lack of oxygen intake[41]. The most common symptoms that manifest during pneumonia are dyspnea, fever, myalgia, cough and cold. Pneumonia can affect one or both lungs and can be caused due to bacterial, viral or fungal invasion and growth in the body [16]. Nosocomial Pneumonia, as the name indicates, is a healthcare-associated illness that was not present at the time of hospitalization and manifests clinically after 48 hours or longer of hospitalization. Pneumonia that originated in a normal environment other than a hospital is known as Community or commonly Acquired Pneumonia (CAP) and is a frequent cause of malaise and loss of life[42]. Common disease-producing pathogens which

can lead to CAP are streptococcus pneumoniae, haemophilus influenzae and moraxella catarrhalis.

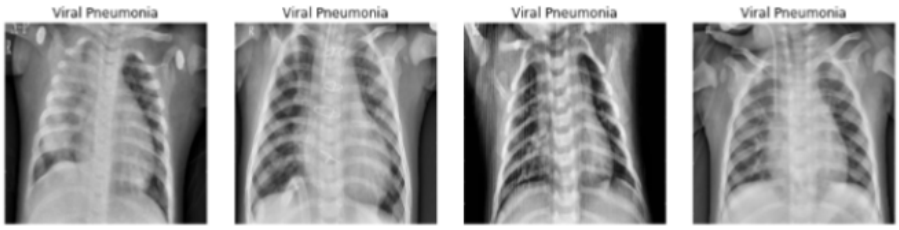

Figure 3: Sample Pneumonia images from C19RD dataset.

1.1.4 Lung Cancer

Lung cancer is a form of cancer that develops in the lungs' tissues, mostly impacting the cells that form the airways[3]. It is one of the most widespread and severe types of cancer, typically resulting in a late-stage diagnosis due to subtle symptoms that encompass a recurrent cough, chest pain, shortness of breath, unexplained weight loss, and exhaustion. However, these symptoms can overlap with other respiratory illnesses such as chronic obstructive pulmonary disease (COPD), pneumonia, or tuberculosis, resulting in misdiagnosis, particularly in the early stages when symptoms are mild. To accurately diagnose lung cancer, a combination of radiological tests is required, with CXRs and CT (Computed Tomography) scans being two of the most commonly utilized imaging modalities.

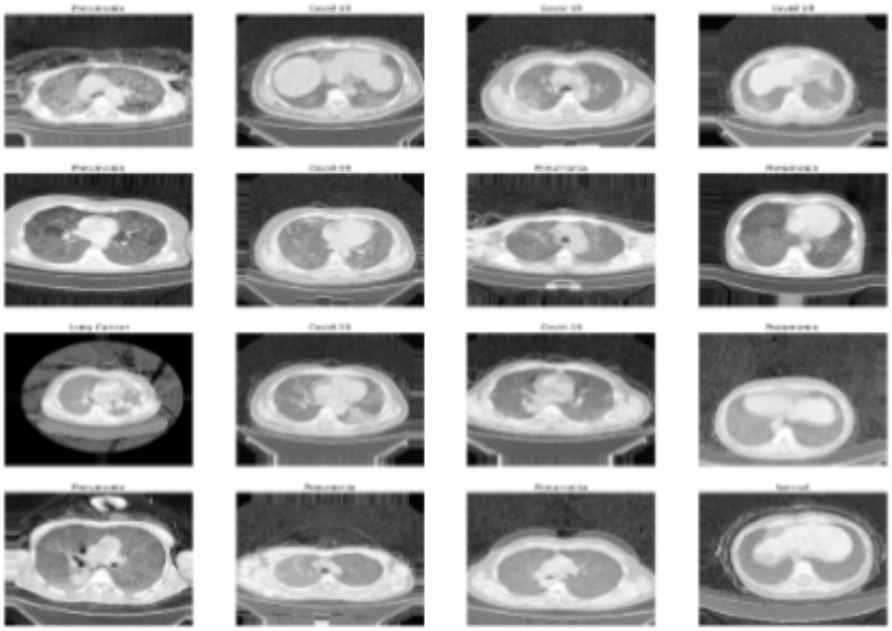

Figure 4: Sample Lung Cancer CT images from IQ-OTH/NCCD and Pneumonia, Normal, COVID-19 CT images from Mendeley datasets.

1.2 Imaging Modalities and Dataset Description

Biomedical devices employing various imaging techniques such as radio waves, electromagnetic waves, magnetic fields, sound waves and such, are used to generate medical image data [27]. Each media mechanism has a distinct methodology and a distinctive response to the human body, organs and tissue structure. The following modalities have been utilized for the research work:

Table 1: The above table lists the datasets that have been utilized for the research work containing details such as the number of classes, number of images per class and the imaging modality.

DATASET NAME	IMAGING MODALITY	NUMBER OF CLASSES	IMAGES PER CLASS
COVID-19 & Common Pneumonia Dataset	CT Scan	2	COVID-19: 722 CAP: 328
SARS-COV-2 Dataset	CT Scan	2	COVID-19: 1252 Non-COVID-19: 1230
Mendeley's COVID Pneumonia Dataset	CT Scan	3	COVID-19: 1680 Pneumonia: 2105 Normal: 2103
Covid-Pneumonia-Normal-chest-x-ray Dataset	X-ray	3	COVID-19: 1525 Pneumonia: 1525 Normal: 1525
COVID-19 Radiography Dataset (Version 2)	X-ray	3	COVID-19: 1143 Pneumonia: 1345 Normal: 1341
COVID-19 Radiography Dataset (Version 5)	X-ray	4	COVID-19: 1345 Pneumonia: 1500 Normal: 1500 Lung Opacity: 1500
(IQ-OTH/NCCD) Lung Cancer Dataset	CT Scan	1	Lung Cancer: 561
COVID_NORMAL_PNEUMONIA _mendeley Dataset	CT Scan	3	COVID-19: 1681 Healthy: 2103 Pneumonia: 2105
Montgomery Dataset	X-ray	1	Tuberculosis: 58
Shenzhen Dataset	X-ray	1	Tuberculosis: 336
Belarus Dataset	X-ray	1	Tuberculosis: 304
NLM Dataset	X-ray	2	Normal: 406 Tuberculosis: 394
RSNA Pneumonia Dataset	X-ray	3	Normal: 3500 Pneumonia: Lung Opacity:

- **Projectional Imaging** : X-rays are a type of electromagnetic waves with specific wavelength which are converted into photo-electrons using the energy level. The operation of image formation is split into three major stages: Image pre-read, main read, and image processing[27],[32]. It is the most economical and commonly used imaging mechanism.

- **Computed Tomography (CT)** : Due to minor variations in attenuation, standard X-ray visualization prediction sometimes neglects to produce good results. CT is commonly used in cancer screening applications such as lung and virtual colonoscopy. Various CT Imaging techniques such as positron emission tomography (PET) , CT angiography and such are available [27],[32]

1.3 Literature Review

The aim of this literature review is to throw light on some of the important works explored by researchers in the area of medical image classification mainly for pulmonary or chest diseases.

The primary goal of the study carried out by Bhatt and Soni(2021) is to provide a straightforward and effective method for finding and obtaining lung CT images from a huge medical image database. In this study, a CBIR-based BOF approach for lung cancer classification is proposed. The approach did not require a comprehensive segmentation of the lung, which sped up the procedure. SURF is a method for finding and describing spatial characteristics in images and discovering regions of interest(ROI) in images of varying sizes, points of view, distances, scale shifts, and rotational invariance, as well as being resistant to other traditional geometric and photometric changes. The graphical patterns were displayed using SURF features and then transformed into groups of distinct pattern categories using K-means. The combination of hierarchical and K-Means clustering yields a more exact and adept image retrieval system and more efficient results than any other clustering technique. With 3,70 lung CT images, the MATLAB image processing toolbox was used to implement the suggested CBIR-based BOF system, which yields 99 % training accuracy and 98.56 % testing accuracy [6].

Bhandary et al.(2020) provided a tool for assessing lung, pneumonia and cancer using Deep Learning (DL). This paper offers two different DL approaches for evaluating the issue in question. The first DL method, known as a modified AlexNet (MAN), is used to categorize CXR images as normal or pneumonia. The MAN categorizes using a Support Vector Machine (SVM), and the results are compared to Softmax. To corroborate its effectiveness, different DL approaches such as AlexNet, VGG16, VGG19, and ResNet50 are used. The second DL study improves lung cancer classification accuracy by combining handcrafted and trained features in the MAN. Serial fusion and Principal Component Analysis (PCA)-based feature extraction are used to improve the feature map. The effectiveness of this DL framework is assessed using LIDC-benchmark IDRI's lung cancer CT images, with a classification accuracy of 97.27% [5].

Houssein et al.(2022) demonstrated a hybrid quantum-classical convolutional neural network (HQ-CNN) model that detects COVID-19 patients using CXR images and random quantum circuits as a foundation. A set of 5445 radiograph visuals was employed to assess the HQ-CNN [19]. The suggested technique could differentiate COVID-19, viral pneumonia, and bacterial pneumonia from normal cases. In an experiment to distinguish between COVID-19 and viral pneumonia cases, it achieved 97.8%, 96.0%, and 96.5% for average accuracy, Kappa score, and MCC, respectively.

Tawsifur Rahman et al.(2021) presented a UNet model for lung segmentation to the conventional UNet model. Six different models, ResNet18, ResNet50, ResNet101, InceptionV3, denseNet201, and ChexNet and a simplistic CNN model were studied on CXR images. The team created a CXR dataset (COVQU) of 18,479 CXR images, including 8851 images showing no pathogen, 6012 non-COVID-19 lung diseases, and 3616 COVID-

19 films. After enhancing the radiographs using various mechanisms, COVID-19 detection was carried out using five different techniques: histogram equalisation, contrast limited adaptive histogram equalisation, image complement, gamma correction, and balance contrast enhancement technique.The U-Net model's accuracy was 98.63%, and classification results from ordinary CXRs were somewhat improved than that from segmented lung CXR images; however, system performance consistency was greatly improved for segmented lung images. The accuracy, precision, sensitivity, F1-score, and specificity for segmented lung visuals were 95.11%, 94.55%, 94.56%, 94.53%, and 95.59%. The proposed method, which is also very reliable, would improve COVID-19 detection by using chest X-ray images [34].

Nkouanga and Vajda's(2021) objective was to develop a single feature descriptor, rather than a mixture of several as advocated and promoted in recent research, that could accurately and concisely explain the various pathological abnormalities in the lungs. Efficiency was measured by utilizing area under the curve (AUC) and accuracy (ACC). AUC for Montgomery samples was 0.81 (ACC = 83.33 percent) and AUC obtained for Shenzhen samples was 0.96 (ACC = 96.35 percent). There are also comparisons to other cutting-edge systems that have already been described in the domain. The fully automatic system that is being proposed processes the CXR image being fed to the model by using image processing methods. This is done to identify the region of interest (ROI) and is followed by performing an economical feature extraction employing edge detection using the Laplacian of Gaussian, which can be enriched by taking into account the local distribution of intensities. Because certain pulmonary abnormalities are concentrated in specific parts of the lungs, the decision to "zoom in" the ROI and check for abnormalities locally was decided. The classifiers are then used to determine whether each lung X-ray is normal or abnormal [30].

Abiyeva et al.(2021) proposed two CNN models that were trained using two separate datasets. With the Kermany et al. dataset, which only comprised pneumonia cases and normal chest X-ray images, the first model was trained for binary classification. Based on the Covid-19 radiography dataset, which comprised of chest X-ray (CXR) images, the second model used transfer learning to build on the knowledge learned by the first model and trained for three class classifications on Covid-19, pneumonia, and normal cases. On the test data, the model performed well in terms of accuracy, recall, precision, and F1 score, with values of 98.3%, 97.9%, 98.3%, and 98.0%, respectively[1].

Xu et al.(2021) proposed a computer-aided diagnostic method that could automatically evaluate CT scans to identify Covid-19 from pneumonia (CAP). The proposed method is based on an unsupervised pulmonary opacity screening approach that unearthed low contrast areas using an unsupervised model conditioned on CT images of normal lung tissues. Inside the opacity zones, radiomics-based characteristics were retrieved and supplied into classifiers for classification. They used 200 CT scans from various patients at several hospitals to assess the proposed CAD system. The accuracy, precision, recall, f1-score, and AUC attained were 95.5%, 100%, 91%, 95.1%, and 95.9% respectively, indicating promising potential on Covid-19 differential diagnosis from CT images [43].

Watershed segmentation for identification and SVM for categorization of detected nodules as malignant or benign were applied in the proposed method given by Makaju et al.(2018). This study was carried out to detect the cancerous nodule from a chest radiograph samples. The proposed model has a 92 % accuracy rate, which is greater than other models, while the classifier has an accuracy rate of 86.6%. In comparison to other models, the suggested system showed overall improvement. However, the proposed

method does not identify between cancer stages I, II, III, and IV. Furthermore, proper pre-processing and the elimination of bogus objects can further improve accuracy [25].

The purpose of the study undertaken by Chowdhury et al.(2020) was to create a reliable method for distinguishing Covid-19 pneumonia from digital CXR imagery. This was carried out by using DL algorithms which have already been trained on another dataset while increasing detection rate. They constructed a public database by integrating multiple public databases and gathering data from recently published studies. A number of DCNNs, already trained using other databases were developed and corroborated using the transfer learning mechanism as well as by creating artificial images by the process of augmentation.The models were trained to classify among the types of pneumonia: i) normal and COVID-19 pneumonia; and ii) normal, viral, and COVID-19 pneumonia with and without augmentating the samples. Both approaches had 99.7% classification accuracy, precision, sensitivity, and specificity of 99.7%, 99.7%, and 99.55% and 97.9%, 97.95%, 97.9%, and 98.8%, respectively[11].

In majority of the research, AI-driven techniques for detecting COVID-19-positive cases are restricted to CT scan or CXR. Using multiple data types to discover abnormal patterns caused by COVID-19 might potentially give more information. As a result, Mukherjee et al.(2021) developed a Convolutional Neural Network (CNN)-tailored Deep Neural Network (DNN) that can train and evaluate both CT scans and CXRs in this research. He attained an overall accuracy of 96.28 % in his studies (AUC = 0.9808 and false negative rate = 0.0208). Furthermore, when combining CT scans and CXRs to detect COVID-19-positive patients, major existing DNNs produced consistent findings[28].

The study carried out by Nahiduzzaman et al.[29] introduced a computer-aided diagnostic framework combining a lightweight parallel CNN for robust feature extraction and an Extreme Learning Machine (CNN-ELM) for classification to identify 17 lung diseases from chest X-rays. The methodology included six experimental trials with varying disease classes (2 to 17), achieving high accuracy (90.92%) and AUC (96.93%) for multiclass classification. It accurately identified COVID-19 (99.37%) and TB (99.98%) in just 0.996 microseconds per image. The framework demonstrated resilience across balanced and unbalanced datasets, binary and multiclass scenarios, and varying image resolutions. A prototype Android app was developed to showcase its potential for real-world deployment.

Shamrat et al.[37] utilized the ChestX-ray14 dataset to classify 14 lung diseases using a fine-tuned MobileLungNetV2 model. Pre-processing included CLAHE for contrast enhancement, Gaussian filtering for noise reduction, and data augmentation. Several transfer learning models were evaluated, with MobileNetV2 performing best (91.6% accuracy), further fine-tuned to achieve 96.97% accuracy. Grad-CAM was employed for visualizing disease detection, and the model demonstrated high precision (96.71%), recall (96.83%), and specificity (99.78%).

Farhan and Yang[14] proposed a Hybrid Deep Learning Algorithm (HDLA) for lung disease classification from CXRs, featuring pre-processing for image quality enhancement, robust 2D CNN-based feature extraction, and optimization using min-max scaling. Extracted features were classified using machine learning models like AdaBoost, SVM, and Random Forest. The approach improved the overall accuracy to 98.35%, an increase of 3.1% and reduces computational complexity by 16.91% compared to state-of-the-art methods.

The proposed DWTMBConvNet model given by Bhosale and Yadav[7] combined discrete wavelet transform (DWT) for frequency-based feature extraction and EfficientNet-inspired MBConv blocks for efficient learning to classify multiple pulmonary diseases, in-

cluding COVID-19 and pneumonia, from chest X-rays. Achieving 95.5% accuracy, 98.9% specificity, and 95.5% sensitivity in five-way classification, it surpassed state-of-the-art methods. Its robust performance across binary, three-way, four-way, and five-way classifications demonstrated its potential as a reliable clinical diagnostic tool.

Ashwini et al.[4] proposed a deep learning-based multi-class classification system using two CNN models—Classification_1 for detecting lung diseases and Classification_2 for classifying specific conditions like COVID-19, pneumonia, TB, and lung cancer. The models utilized grid search optimization (GSO) for hyper-parameter tuning, achieving 99.82% and 98.75% accuracy, respectively, on a substantial dataset of CXR images. The results outperformed existing methods, offering a reliable tool to aid medical professionals in rapid and accurate lung disease diagnosis. Future work could explore integrating advanced architectures, multi-modal imaging data, and real-time deployment to enhance model performance and scalability in clinical settings.

The next study proposed a hybrid framework for respiratory lung disease detection, combining a custom convolutional neural network (CCNN) for feature extraction with quantum classifiers (MMS and MSMS) based on quantum variational circuits. The model was tested on the COVID-19 Radiography Dataset (CRD) with 15,153 chest X-ray images, achieving training and testing accuracies of 98.9% and 98.1%, respectively, with low computation costs. The proposed hybrid model outperformed standard deep learning models and was validated using real-time deployment on the IBM Q-QASM quantum computer[35].

Cutia et al.[13] proposed a method for lung disease detection using CXRs, applying the Contrast-Limited Adaptive Histogram Equalization (CLAHE) technique and enhancing DenseNet201 with hybrid pooling and channel attention mechanisms. The model outperformed several well-known pre-trained models, achieving accuracy, precision, recall, and F1-scores of 95.34%, 97%, 96%, and 96%, respectively. Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to provide visual insights into the model's decision-making, aiding radiologists in diagnosing conditions like pneumothorax and atelectasis. Future work could extend the model to other diseases with X-ray, CT scans, and MRIs, and explore explainability models like saliency maps. Testing on diverse medical and non-medical datasets could provide further insights.

The study carried out by Indumathi and Siva[20] put forward a hybrid Crystal-optimized BiDLSTM-Mask-RCNN model for detecting lung diseases using CXRs. The model was evaluated on the COVID-19 radiography dataset, Tuberculosis (TB) Chest X-ray Database, and the National Institute of Health Chest X-ray Dataset. The results show that the proposed model outperformed conventional methods with an accuracy of 99%, precision of 97%, recall of 96%, F1-score of 96%, and specificity of 98%. The optimized Mask-RCNN model achieved a logistic value of 0.99, demonstrating its superior performance in lung disease detection. Additionally, the model achieved higher testing and training accuracy and lower testing and training loss compared to traditional models. Future work would involve identifying the severity of COVID-19 using multi-modality imaging to further enhance diagnosis.

Karaddi and Sharma proposed the use of eight pre-trained convolutional neural networks (CNNs) to classify lung diseases, including COVID-19, pneumonia, pneumothorax, tuberculosis, and normal conditions, from CXRs. The classification process involved two phases: training the networks with the Adam optimizer and using the trained models for classification. The dataset underwent preprocessing, resizing, and data augmentation. Among the eight networks, DenseNet-201 achieved the highest performance, with accu-

racy, sensitivity, and specificity values of 97.2%, 94.28%, and 97.92% for K=5, respectively, and further improvements for K=10 and K=15. The proposed method outperformed existing state-of-the-art approaches and could assist clinicians in making faster diagnoses for better treatment outcomes[21].

Bhosale and Patnaik[8] proposed a unique CNN model, PulDi-COVID, for detecting nine pulmonary diseases, including COVID-19, using chest X-ray images (CXI). The model addressed challenges like noisy and small data by employing an SSE ensemble learning approach, combining multiple deep learning models for more accurate predictions. PulDi-COVID was compared with thirteen existing studies and achieved impressive results, including 99.70% accuracy, 98.68% precision, 98.67% recall, 99.24% AUC-ROC score, and a low error rate of 1.33%. These findings demonstrate that PulDi-COVID outperforms current state-of-the-art methods in COVID-19 and chronic lung disease detection, offering a reliable tool for early diagnosis and improving patient outcomes.

1.4 Challenges in Pulmonary Disease Classification

Several challenges were identified in the field of pulmonary disease classification. They are as follows:

- How to extract relevant features from a scanty medical image dataset? Because medical images are generally difficult to obtain, medical datasets are usually small, and we cannot retrieve discriminant features from them. Even if the proposed model achieves very high classification accuracy irrespective of the size of the image dataset, the specific application value is severely limited[24].

- Unbalanced datasets: It is to be mentioned that the majority of the datasets utilized for binary or multi-class classification in pulmonary disease detection are highly skewed. The lopsidedness of the dataset can lead to bias in a model's efficiency[2].

- Quality of Images: Efforts are being made to enhance the contrast of the medical images as they are generally low contrast. The diverseness in the quality of images collected from different places leads to bias [2].

- Distinction from other Viral diseases: One of the significant challenges in pulmonary disease classification is accurately distinguishing one disease from another, as many pulmonary conditions share overlapping radiographic patterns and clinical features. For example, diseases such as viral pneumonia, bacterial pneumonia, and COVID-19 may all present with similar opacities, consolidations, or ground-glass appearances on chest X-rays or CT scans. These shared characteristics can make it difficult for both traditional and AI-based methods to differentiate between them, especially in the absence of additional clinical or pathological context.

- Generalization: The potential of a learning algorithm to perform effectively on an unknown dataset is referred to as generalization. This necessitates training the model on a diversified dataset. Some strategies that can be used to improve generalization skills include regularization, dropout, batch normalization, and early stopping [2].

- How to quickly and efficiently fuse different types of features from different models? To overcome this challenge we need to design a more satisfactory fusion approach so that all the feature vectors can be combined into a larger vector. This will enable us to determine one proportion parameter between different features [2].

- High correlation of errors when using ensemble techniques: It is expected that the base classifiers used in the ensemble should complement each other to produce better classification results. Most of the studies surveyed only combine classifiers that were trained on similar features. This causes the correlation error of the base classifiers to be high [23].

2 Objectives

After an extensive review of the available literature and the previously done research work, we have come across certain gaps in areas that we will try to overcome. Below given are some of the objectives that are in our focus :

1. To develop newer feature extraction methods that achieve better classification accuracy for pulmonary diseases.

2. To develop a robust model that can distinguish between various pulmonary diseases such as Covid-19, lung cancer, viral pneumonia even for unbalanced datasets.

3. To explore Transfer Learning techniques for developing pulmonary/lung disease classification models that give better and more accurate results.

4. Designing of Ensemble learning approaches for multi-class pulmonary disease classification.

3 Proposed Method 1: Hand-crafted Textural and Deep Features for Multi-Class Lung Disease Classification

3.1 Methodology

Pulmonary diseases present a significant challenge in medical diagnostics due to their diverse etiologies and overlapping clinical symptoms, necessitating advanced classification methods to enhance diagnostic accuracy and treatment efficacy. This study focuses on multi-class lung disease classification using CT scans, leveraging advanced texture descriptors including Gabor filters, Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG) for feature extraction. These features are integrated with the DenseNet121 architecture, which is also employed for extracting deep features from CT images. Various ML algorithms, such as Random Forest (RF), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Decision Tree (DT), Naive Bayes (NB), and K-Nearest Neighbors (KNN) are employed for classification.

The integration of CT imaging with ML and DL algorithms significantly enhances diagnostic accuracy of pulmonary diseases by meticulously analyzing medical images. This study proposes a multi-class pulmonary disease classification model as shown in Fig.5, capable of classifying input images into four distinct target classes, thereby improving diagnostic precision and facilitating timely and appropriate medical intervention. Before proceeding with feature extraction and classification, the CT images were converted to grayscale, resized to 224 * 224 dimension and normalized. Gamma correction and histogram equalization are two techniques that were used to enhance the quality of CT scan images. Gamma correction adjusted image brightness using a nonlinear transformation,

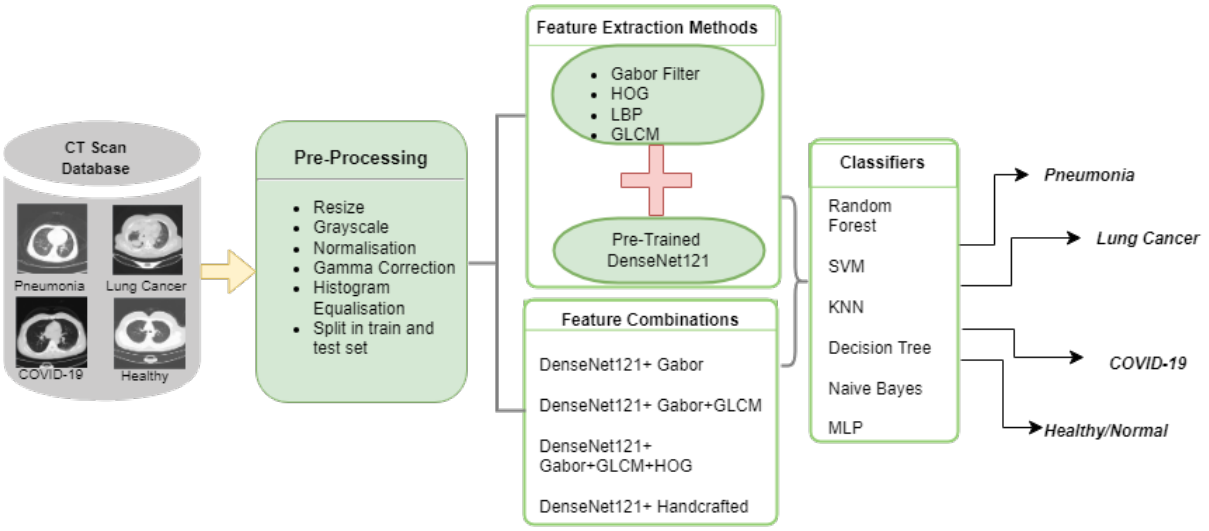

Figure 5: Block Diagram of the Proposed Model.

which enhanced the visibility of details. The transformation can be defined as $I_{\text{new}} = I^{\gamma}$, where I is the normalized pixel value and γ is the gamma value. For $\gamma < 1$, the image lightens, and for $\gamma > 1$, it darkens. On the other hand, histogram equalization improved contrast by redistributing pixel intensity values to utilize the full intensity range. The pre-processing steps standardized and enhanced input image data, optimizing it for machine learning models.

3.2 Results

All classification models were run on a system equipped with an Intel® Core™ i7-1065G7 processor, operating at 1.30GHz to 1.50GHz, along with 16 GB of RAM and a 4 GB GPU, all under a 64-bit Windows 11 operating system. ImageDataGenerator module in Keras was employed for data augmentation. Dataset images were split in an 80:20 ratio for training and validation. The train set contained 5160 CT images whereas 1290 images were present in the Validation set. Nine different feature combinations along with six classifiers were put to use in the research study, taking the total number of experiments to 54.

In the research study, various feature extraction techniques as shown in Table2, have been employed to derive relevant features from CT scan images, which are essential for the accurate prediction of lung diseases such as TB, Pneumonia, COVID-19, and Lung Cancer. Features are measurable attributes or properties extracted from images that capture key information about the underlying structures and patterns. These attributes include texture, shape, and intensity variations within the images. By employing different feature extraction methods, the study aims to highlight and quantify these important characteristics, enabling more precise and effective classification of lung diseases. For classification purposes, our study utilized several ML algorithms, including SVM, Random Forest, Decision Tree, MLP, Naive Bayes and KNN.

Table 2: Feature combinations fed to the individual classifiers.

Feature Extraction Technique	**No. of Features**	**Feature Combinations**
LBP	10	Feat_Set01
HOG	26244	Feat_Set02
Gabor	128	Feat_Set03
GLCM	25	-
DenseNet121	50176	Feat_Set04
Handcrafted (LBP+HOG+Gabor+GLCM)	26407	Feat_Set05
DenseNet121+GLCM	50201	Feat_Set06
DenseNet121+(GLCM+Gabor)	50329	Feat_Set07
DenseNet121+(GLCM+Gabor +HOG)	76573	Feat_Set08
DenseNet121+Handcrafted	76583	Feat_Set09

3.3 Discussion

The results obtained after training and testing the models with different sets of features are listed in Table 3 and 4. Analyzing the performance of various classifiers reveals significant insights across feature sets. The Support Vector Machine (SVM) classifier shows

Table 3: Result for models applying individual handcrafted and deep feature extractor.

Feature Extraction Technique	**Classifiers**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Feat_Set01 (LBP)	RF	94.72	95.66	95.80	95.73
	SVM	41.62	45.04	38.75	38.36
	KNN	95.65	96.53	96.55	96.52
	NB	50.85	54.09	55.70	50.27
	DT	87.51	89.68	89.80	89.73
	MLP	64.41	65.45	64.97	64.43
Feat_Set02 (HOG)	RF	97.36	98.04	97.48	97.73
	SVM	99.68	99.77	99.76	99.76
	KNN	99.84	99.88	99.73	99.80
	NB	80.46	79.82	83.09	81.00
	DT	89.14	89.74	89.31	89.51
	MLP	99.53	99.65	99.62	99.64
Feat_Set03 (Gabor)	RF	98.21	98.47	98.41	98.44
	SVM	82.94	86.50	85.22	85.83
	KNN	98.91	99.11	99.15	99.12
	NB	48.06	50.58	57.31	46.10
	DT	92.48	92.29	92.94	92.94
	MLP	83.87	87.85	86.07	86.54
Feat_Set04 (DenseNet121)	RF	98.21	98.66	96.99	97.74
	SVM	99.92	99.94	99.94	99.94
	KNN	99.84	99.88	99.72	99.80
	NB	84.34	87.19	84.98	85.84
	DT	89.68	89.19	89.85	89.61
	MLP	99.76	99.82	99.80	99.81

Table 4: Result for models applying various combinations of handcrafted and deep feature extractors.

Feature Extraction Combination	**Classifiers**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Feat_Set05	RF	97.90	98.27	97.98	98.11
	SVM	99.68	99.77	99.76	99.76
	KNN	99.84	99.88	99.73	99.80
	NB	80.85	80.36	83.39	81.41
	DT	91.00	91.04	91.36	91.18
	MLP	99.45	99.59	99.54	99.57
Feat_Set06	RF	98.60	98.97	97.45	98.17
	SVM	99.92	99.94	99.94	99.94
	KNN	99.84	99.88	99.72	99.80
	NB	84.34	87.19	84.98	85.84
	DT	89.92	90.16	90.10	90.13
	MLP	99.76	99.82	99.80	99.81
Feat_Set07	RF	98.52	98.91	97.52	98.18
	SVM	99.92	99.94	99.94	99.94
	KNN	99.84	99.88	99.72	99.80
	NB	84.41	87.24	85.04	85.91
	DT	89.30	89.47	89.03	89.24
	MLP	99.84	99.88	99.86	99.87
Feat_Set08	RF	99.14	99.36	98.94	99.15
	SVM	99.92	99.94	99.94	99.94
	KNN	99.84	99.88	99.72	99.80
	NB	86.51	88.47	88.01	88.14
	DT	88.60	88.56	88.17	88.34
	MLP	99.84	99.88	99.86	99.87
Feat_Set09	RF	99.76	99.82	99.82	99.82
	SVM	99.84	99.88	99.88	99.88
	KNN	99.68	99.76	99.59	99.67
	NB	89.37	90.42	89.63	89.97
	DT	94.10	94.80	94.56	94.67
	MLP	99.84	99.88	99.88	99.88

a broad range of performance, with a low accuracy of 41.63% in Feat_Set01 but achieving near-perfect scores in Feat_Sets04, 06, 07, and 08, indicating its high potential when optimal features are used. K-Nearest Neighbors (KNN) consistently performs well, particularly excelling in Feat_Sets02 and 09, with accuracies ranging from 95.66% to 99.84%, demonstrating its reliability. The Multi-Layer Perceptron (MLP) also exhibits strong performance, especially in Feat_Sets02 through 09, consistently achieving accuracies above 99

In contrast, Random Forest (RF) and Decision Tree (DT) classifiers show moderate but reliable performance, with RF achieving its highest accuracy in Feat_Set09 (99.77%). RF's performance ranges from 94.73% to 99.77%, showing consistency and robustness. Decision Tree's performance is more variable, ranging from 87.52% in Feat_Set01 to 94.11% in Feat_Set09. Naive Bayes (NB) exhibits significant variability, with performance ranging from 48.06% in Feat_Set03 to 89.38% in Feat_Set09, indicating less reliability compared to other classifiers.

The combination of DenseNet121 features with other handcrafted features notably enhances classifier performance, particularly for SVM, KNN, and MLP. These classifiers emerge as the most effective for this dataset, achieving high accuracy and consistency. Overall, this study underscores the importance of selecting and combining appropriate feature extraction techniques to optimize classifier performance in medical image classification tasks.

Despite slightly lower performance metrics for SVM and KNN in comparison to other feature sets, Feat_Set09 still demonstrated strong results with accuracies of 99.84% and 99.68% respectively. Overall, Feat_Set09 proves to be the most effective feature combination for robust and accurate classification of pulmonary diseases.

4 Proposed Method 2: Differential Diagnosis of Pulmonary Diseases using Convolutional Neural Network and its variants

4.1 Convolutional Neural Network with LSTM Architecture

4.1.1 Methodology

Pulmonary disorders, such as pneumonia, tuberculosis, and COVID-19, have a significant global health impact, requiring reliable and swift diagnostic methods to improve treatment outcomes. This study examines the potential of a hybrid CNN-LSTM model to surpass conventional methods in accurately diagnosing various lung diseases from chest X-rays (CXRs). The study has been carried out to successfully categorize lung diseases into 4 target classes: Pneumonia, Tuberculosis, COVID-19 and Normal. The proposed CNN-LSTM model leverages CNN for feature extraction from 2D image data, capturing detailed spatial patterns relevant to pulmonary diseases. The LSTM component in this model is employed to integrate contextual information across image patches. This modification allows the model to effectively capture spatial relationships between different regions of CXR images, enhancing its ability to detect subtle abnormalities and improving overall diagnostic accuracy in pulmonary issues. The proposed model has further been compared with baseline CNN and CNN-Bi LSTM models to prove its superior performance. The methodology employed in this study, as depicted in Figure 6 comprises of i) Data collection, ii) Data Pre-processing that includes image enhancement and denoising, iii) Training and testing the proposed CNN-LSTM model, iv) Training and testing CNN and BiLSTM

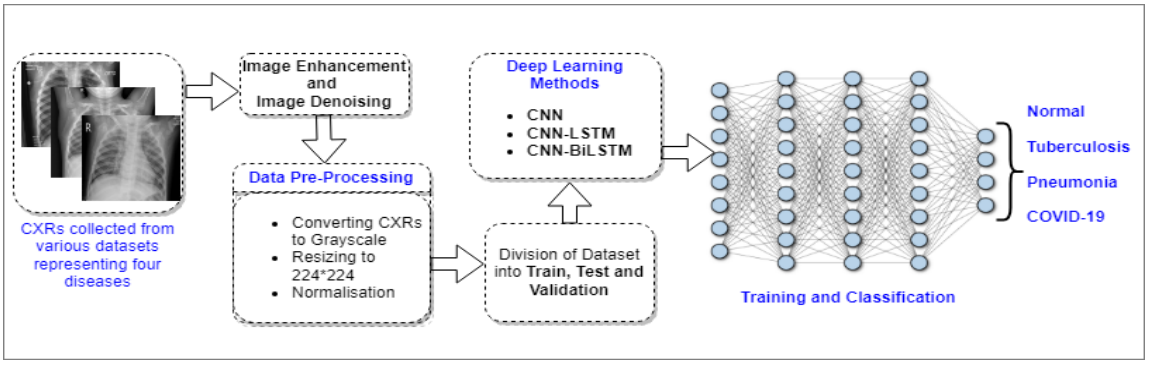

Figure 6: Architecture of the Proposed Model.

models and v) Comparative analysis to assess model performance. The results have been listed in subsection 4.2.2.

4.1.2 Results

The evaluation results as shown in Table 5, reveal notable differences among the three models used for pulmonary disease classification. The basic CNN model achieved a loss of 0.2215, indicating its ability to minimize errors during training. It exhibited an accuracy of 94.26%, implying that it correctly classified nearly 94.26% of instances. On the other hand, the proposed CNN-LSTM model demonstrated superior performance across all metrics. With a lower loss of 0.1473, it exhibited higher accuracy, precision, recall, and F1 Score compared to the basic CNN model. The accuracy of 96.24% indicates its ability to classify instances with greater accuracy, while precision and recall values of 96.44% and 96.62%, respectively, signify its improved capability in correctly identifying positive cases and capturing actual positive instances. The F1 Score of 96.49% further highlights the balanced performance of the CNN-LSTM model.

Table 5: Results obtained for proposed models.

Models	**Loss**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
CNN	0.2215	94.26	93.84	93.94	93.85
CNN-LSTM	0.1473	96.24	96.44	96.62	96.49
CNN-Bi LSTM	0.3183	95.58	95.31	94.94	95.05

Table 6: Classification Report for CNN model

Class	**Precision**	**Recall**	**F1 Score**	**Support**
Normal	0.94	0.99	0.96	136
Tuberculosis	0.91	0.92	0.91	74
COVID-19	0.93	0.93	0.93	107
Pneumonia	0.98	0.92	0.95	136

Table 7: Classification Report for **proposed CNN-LSTM** model

Class	**Precision**	**Recall**	**F1 Score**	**Support**
Normal	0.93	0.98	0.95	136
Tuberculosis	0.96	0.99	0.97	74
COVID-19	0.99	0.98	0.99	107
Pneumonia	0.98	0.92	0.95	136

Table 8: Classification Report for CNN-Bi LSTM model

Class	**Precision**	**Recall**	**F1 Score**	**Support**
Normal	0.96	0.99	0.97	136
Tuberculosis	0.94	0.89	0.92	74
COVID-19	0.91	0.97	0.94	107
Pneumonia	1.00	0.94	0.97	136

Similarly, the CNN-Bi LSTM model showcased competitive performance, albeit slightly lower than CNN-LSTM. With a loss of 0.3183, it achieved an accuracy of 95.58% and demonstrated precision, recall, and F1 Score values of 95.31%, 94.94%, and 95.05%, respectively. These metrics collectively illustrate the effectiveness of incorporating LSTM layers into the CNN architecture, particularly in enhancing accuracy, precision, and recall in classifying lung diseases. The model correctly classified 133 instances as normal (true negatives). However, there were three instances of normal misclassified as viral pneumonia (false positive). All 73 instances of TB were correctly classified (true positives). There was one instance of TB misclassified as COVID-19 (false negative). All 105 instances of COVID-19 were correctly classified (true positives). However, there were two instances of COVID-19 misclassified as TB (false positive). There were 125 instances of viral pneumonia correctly classified (true positives). However, there were ten instances misclassified as normal (false negative), and one instance misclassified as TB (false positive). Similarly, in the case of CNN, the model accurately classified most instances for each class. However, there were some misclassifications: one normal instance was classified as TB and one as COVID-19. Additionally, five TB instances were misclassified as COVID-19, and one as viral pneumonia. Seven COVID-19 instances were mistakenly classified as TB. Furthermore, nine viral pneumonia instances were misclassified as normal, and two as TB. For CNN-Bi LSTM the model accurately classified most instances for each class. However, there were some misclassifications: one instance each of normal and viral pneumonia was misclassified as TB. Moreover, eight TB instances were mistakenly classified as COVID-19. Additionally, three instances of COVID-19 were misclassified as TB. Lastly, six instances of viral pneumonia were mistakenly classified as normal, and two as TB. Table 6, 7 and 8 provide the classification details for each class.

4.1.3 Discussion

In this research, we proposed the CNN-LSTM model for enhancing the classification of multi-class pulmonary diseases. Alongside, we utilized two other baseline models, CNN and CNN-Bi LSTM, for comparative analysis. The hybrid CNN-LSTM model emerged as

the most effective, demonstrating robust performance in disease classification with accuracy, precision, recall, and F1 Score of 96.24%, 96.44%, 96.62%, and 96.40%, respectively. Our experimental evaluation showcased improved accuracy compared to existing state-of-the-art models, validating our model's capability to achieve well-balanced results in terms of AUC-ROC and F1 scores. Moving forward, our future investigations will focus on enhancing the model's generalization by leveraging larger datasets and expanding the scope of classified lung diseases. Additionally, addressing the challenge of class imbalance will remain a key area of focus in our forthcoming studies.

4.2 CNN-GRU Model for Tuberculosis Detection Enhanced with LIME Interpretability

4.2.1 Methodology

Annually, 10 million individuals are affected by Mycobacterium tuberculosis, leading to 1.5 million fatalities. There is an acute requirement to implement plan of action to strengthen the diagnosis and initiation of care for people with TB. This has been the motive behind carrying out this research study. This research study investigates the fusion of a traditional Convolutional Neural Network (CNN) with a Recurrent Neural Network (RNN), specifically the Gated Recurrent Unit (GRU), to improve the categorization of CXRs into tuberculosis and normal categories. The approach used in this study, as shown in Figure 7, includes the following steps: i) Pre-processing, ii) Augmentation of minority class CXR images, iii) Image Enhancement, iv) Training and testing the proposed CNN-GRU model v) Application of LIME(Local Interpretable Model-agnostic Explanations) to identify areas of the image that contribute most to the classification and vi) Analysing the result listed in Table 9 and performing a comparative study with SOTA works.

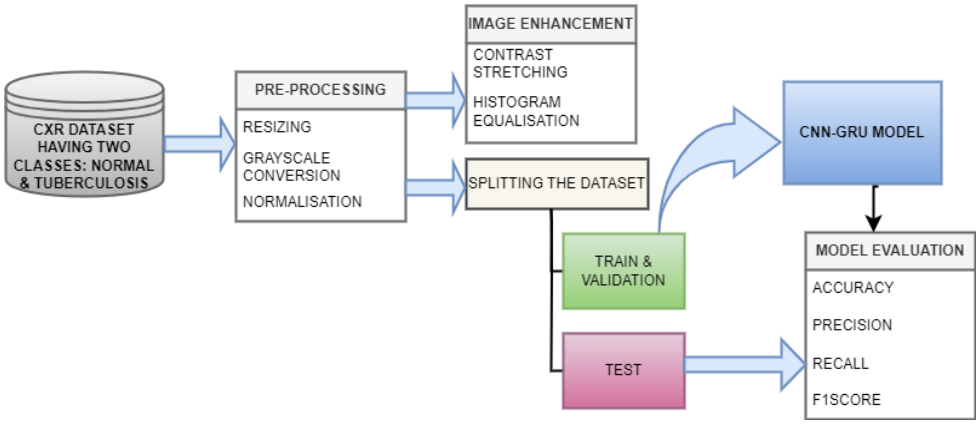

Figure 7: Block diagram of the proposed model.

4.2.2 Results

Binary classification was performed in the given study to obtain accurate and reliable predictions for TB-infected and normal classes. The proposed model was evaluated on training, validation and testing subsets. It achieved an accuracy of 99.96% on the train subset and 98.9% for the validation subset. Upon being tested for unseen test data, it achieved 98.00% accuracy. The confusion matrix reveals that the classification model distinguished between normal and TB-infected cases with high accuracy. The matrix showed

that out of 350 normal images, 348 were correctly classified, while 2 were misclassified as TB-infected. Conversely, out of 350 TB infected images, 338 were correctly identified, with only 12 misclassified as normal. This results in an overall accuracy of 98%, indicating that the model correctly predicts a large majority of the cases. The precision for TB-infected cases is 99.43%, meaning that most of the images predicted as TB-infected are indeed TB-infected. The recall for TB-infected cases is 96.67%, showing that the model successfully identifies most of the actual TB-infected images. The F1 score, combining precision and recall, is 98.02%, reflecting a balanced performance in detecting TB-infected cases with minimal errors. The accuracy and loss curve depicts that the model is learning well on the training subset. To avoid overfitting of the model it has been trained for 50 epochs only. The train and validation loss of the proposed model stands at 0.008 and 0.03 respectively.

Table 9: Result obtained by the proposed CNN-GRU model.

SUBSET	ACCURACY	PRECISION	RECALL	F1 SCORE	TIME(in secs)
TRAINING	99.96	99.96	99.96	99.96	3438.30
TESTING	**98.00**	**99.43**	**96.67**	**98.02**	506.2

LIME for interpretability

LIME (Local Interpretable Model-agnostic Explanations) is a strategy for understanding and interpreting DL models' predictions [38]. In our work, LIME contributes to TB detection, as shown in Figure 8 and 9. LIME utilizes superpixels for pertubation. Initially, CXR images are pre-processed and segmented into superpixels. This segmentation simplifies the image while preserving essential features. LIME perturbs the segmented superpixels, generating perturbed images with modified regions. These images are passed through the model, and their predictions are recorded. By analyzing the predictions of the perturbed images, LIME identifies the superpixels most influential in the model's decision for each class (positive or negative). These identified superpixels represent the regions of the CXR image most relevant for TB detection, providing insights into the model's decision-making process.

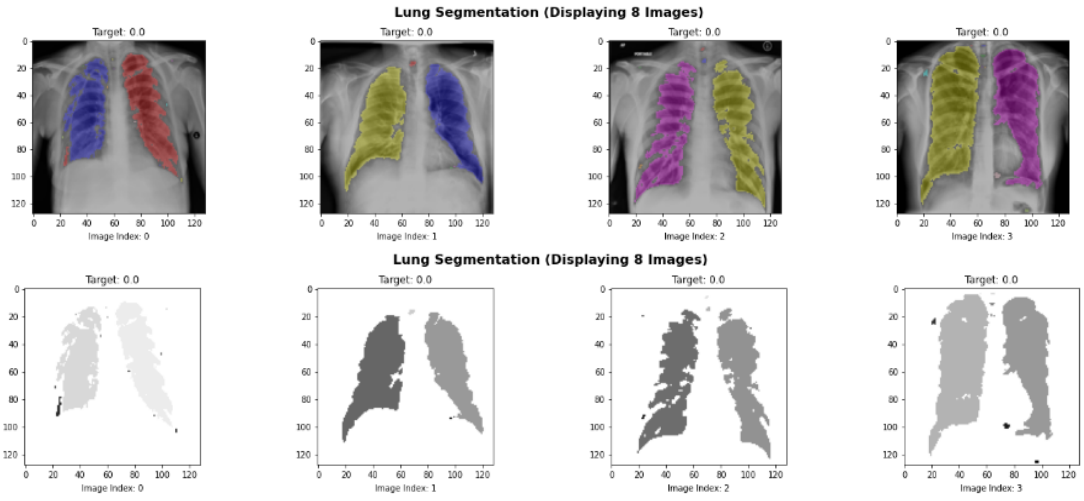

Figure 8: Segmented areas of input CXRs.

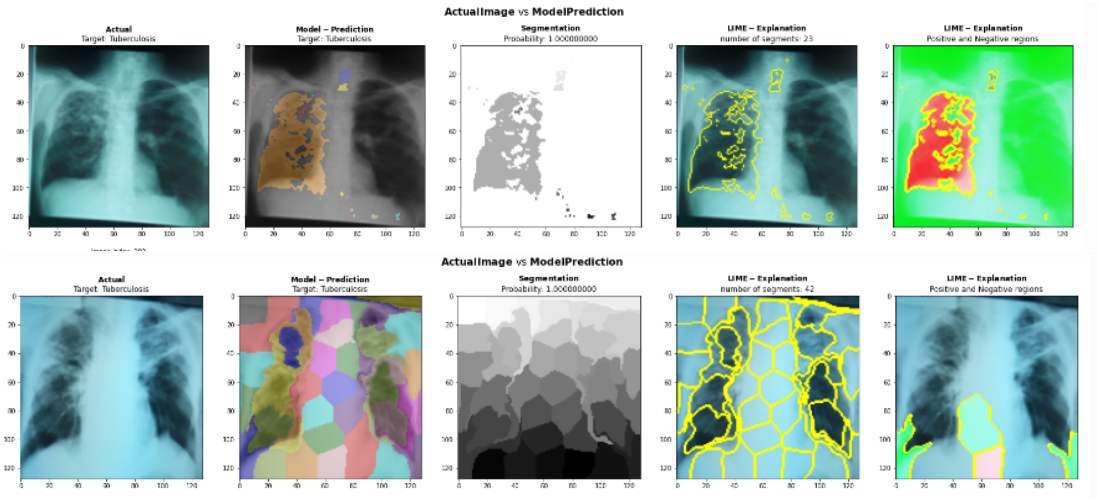

Figure 9: Actual and predicted images with segmentation, LIME explanation and LIME positive and negative regions.

4.2.3 Discussion

The research presents a promising advancement in the field of automated tuberculosis (TB) detection using chest X-rays. The integration of CNN and GRU combines the strengths of both architectures. The CNN excels at feature extraction from raw pixel data, while the GRU specializes in sequential analysis and capturing temporal dependencies. By leveraging the feature representations learned by the CNN and further refining them through sequential processing by the GRU, the model can achieve enhanced discriminative power and classification performance. Data augmentation through transformations such as rotation, scaling, and flipping improved the model's robustness and generalization ability by exposing it to a wider variety of image variations. Moreover, contrast stretching and histogram equalization, enhanced the quality of input images, aiding in the extraction of relevant features and improving the model's performance. In real-world settings, the model could significantly improve TB detection, particularly in resource-limited areas, by aiding early diagnosis and treatment, ultimately impacting global TB control efforts. In communities with a high TB incidence and limited access to qualified radiologists, this model might be incorporated into mobile health units or telemedicine systems, allowing for remote TB screening and diagnosis. This would allow for more timely intervention, especially in underserved communities, potentially lowering TB death and morbidity rates. Furthermore, the LIME package's ability to highlight significant features utilized in predictions provides the model with transparency and interpretability, both of which are critical for garnering the trust of healthcare providers and patients. While the results are encouraging, there remain avenues for future exploration and refinement. Future work could involve the utilization of a more diverse and extensive dataset to enhance the robustness and generalizability of the model. Additionally, testing the model with different optimizers could provide insights into improving its efficiency and effectiveness. Multi-class classification, advanced architectures, and multi-modal data integration could be looked into for further enhancing model performance. Overall, our study contributes to the ongoing efforts to develop sophisticated AI-driven solutions for TB detection, with the ultimate goal of enhancing healthcare outcomes and addressing public health challenges.

4.3 Customized CNN and Oversampling Techniques: A Deep Learning Approach

4.3.1 Methodology

Rapid, sensitive and accurate tests are critical for contemporary health care. Tuberculosis has been with us since time immemorial, and although being treatable and curable, it remains one of the world's most infectious malady, killing millions each year. Significant advances have been made in tuberculosis diagnostic technologies however, they are not adequately portable, rendering them inapplicable in remote, rural areas where they are most required. They are costly, lack the essential sensitivity or precision, and are frequently susceptible to error. In this regard, an automated system that can analyze and diagnose pulmonary TB by using Chest X-Rays(CXR) can prove to be a cost-effective, dependable and quick alternative to manual diagnostic methods.

The research work deals with developing a computer-assisted model, as depicted in Fig 10, that can identify TB from chest radiographs taken for patients. Although TB is a curable and preventable ailment, still its eradication has proved to be difficult. Covid-19 further exacerbated the spread and containment of TB since many cases went undetected during the pandemic. Existing patients too, were unable to access the medical treatment essential for them, worsening their health conditions and impeding their recovery process. The proposed model can help identify and isolate patients suffering from this malady, thereby aiding in the control of its transmission. The methodology employed in this

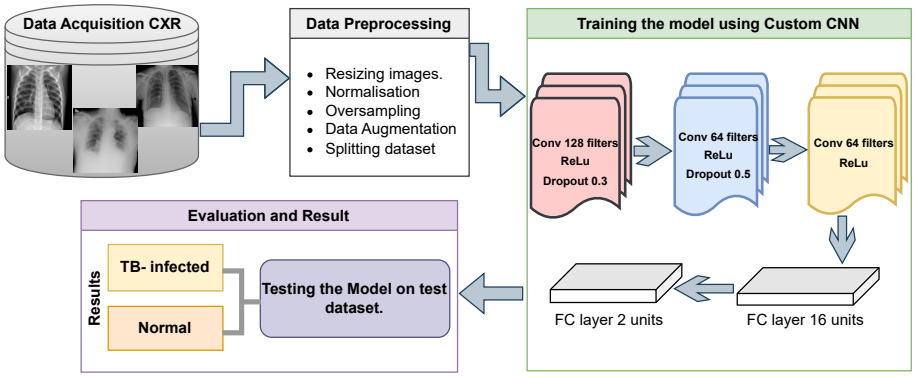

Figure 10: Block diagram of the proposed methodology.

work contains the following steps: i) Data Acquisition and pre-processing, ii) Applying oversampling techniques to overcome the issue of imbalanced target classes, iii) Training the proposed model and classification using custom CNN iv) Comparative study using DenseNet169 and ResNet50, v) Analysis of result and computation time for each classifier and v) Using GRAD-CAM technique to highlight the areas of concern in the CXRs.

4.3.2 Results

Binary classification was carried out in this research work by using different combinations of oversampling mechanisms and DL classifiers. The final outputs obtained thereafter have been listed in Table 10, 11 and 12.

Table 10 shows the results of a Custom CNN model with different oversampling techniques: W/O SMOTE, With SMOTE, Weighted Averaging, ADASYN, and Borderline SMOTE. The models were evaluated on training, validation, and testing datasets. The

Table 10: Results obtained by **Proposed Custom CNN** and various oversampling techniques.

Model	Sampling Method	Dataset	Accuracy	Precision	Recall	F1-Score
Custom CNN	W/O SMOTE	Training	99.97	99.80	100	99.90
		Validation	97.92	95.31	93.85	94.57
		Testing	97.74	91.55	94.89	93.19
	With SMOTE	Training	99.98	99.96	100	99.98
		Validation	98.21	97.89	98.59	98.24
		Testing	99.14	98.43	99.85	99.14
	Weighted Averaging	Training	99.97	100	99.80	99.90
		Validation	99.11	98.44	96.92	97.67
		Testing	98.93	96.38	97.08	96.73
	ADASYN	Training	100	100	100	100
		Validation	99.82	99.64	100	99.82
		Testing	**99.79**	**99.71**	**99.85**	**99.78**
	Borderline SMOTE	Training	99.96	100	99.92	99.96
		Validation	100	100	100	100
		Testing	99.57	99.71	99.42	99.56

oversampling techniques improved accuracy and performance across the board. ADASYN and Borderline SMOTE achieved perfect or near-perfect accuracy, precision, recall, and F1-Score. SMOTE and Weighted Averaging also yielded notable improvements. These techniques effectively addressed class imbalance and enhanced the model's classification performance.

The DenseNet169 model achieved high accuracy with all oversampling techniques: W/O SMOTE, With SMOTE, Weighted Averaging, ADASYN, and Borderline SMOTE. With SMOTE and ADASYN, the model achieved perfect or near-perfect accuracy, precision, recall, and F1-Score. Weighted Averaging and Borderline SMOTE also yielded impressive results. Table 12 showcases the results obtained by the ResNet50 model. The model's performance varied across the techniques and datasets. Without SMOTE, the model achieved high accuracy on the training dataset (94.25%), but lower precision, recall, and F1-Score on the validation and testing datasets. With SMOTE, the model showed improved performance with balanced precision, recall, and F1-Score across all datasets. Weighted Averaging led to a decrease in performance, particularly in recall, while ADASYN and Borderline SMOTE resulted in relatively balanced performance across the metrics. Overall, the results highlight the impact of oversampling techniques on the ResNet50 model's performance, emphasizing the importance of selecting an appropriate technique for addressing class imbalance.

As is evident from the results table, the CNN+ADASYN model and Densenet169 + ADASYN model are at par with each other in accuracy, precision, recall and F1-Score that stands at 99.79%, 99.71%, 99.85% and 99.78% respectively. The training and testing metrics are same for each model but in the case of validation, CNN model performs better than Densenet169 with 99.82%, 99.64%, 100% and 99.82%. The second highest achiever has been CNN+BorderlineSMOTE with 99.57% and DenseNet169 + SMOTE with 99.71% accuracy. The ResNet50 model performs reasonably well along with SMOTE,

Table 11: Results obtained by DenseNet169 and various oversampling techniques.

Model	Sampling Method	Dataset	Accuracy	Precision	Recall	F1-Score
DenseNet169	W/O SMOTE	Training	99.97	100	99.80	99.90
		Validation	98.51	100	92.31	96
		Testing	98.93	97.76	95.62	96.68
	With SMOTE	Training	100	100	100	100
		Validation	99.69	99.30	100	99.65
		Testing	99.71	99.71	99.71	99.71
	Weighted Averaging	Training	100	100	100	100
		Validation	99.11	100	95.38	97.64
		Testing	99.29	97.81	97.81	97.81
	ADASYN	Training	100	100	100	100
		Validation	99.64	99.28	100	99.64
		Testing	**99.79**	**99.71**	**99.85**	**99.78**
	Borderline SMOTE	Training	99.96	99.96	99.96	99.96
		Validation	99.46	99.30	99.65	99.47
		Testing	99.64	99.57	99.71	99.64

Table 12: Results obtained by ResNet50 and various oversampling techniques.

Model	Sampling Method	Dataset	Accuracy	Precision	Recall	F1-Score
ResNet50	**W/O SMOTE**	Training	94.25	93.55	69.88	80.0
		Validation	95.54	93.1	83.08	87.8
		Testing	**93.21**	**87.04**	**68.61**	**76.73**
	With SMOTE	Training	92.42	93.55	90.43	92.29
		Validation	90.81	91.3	89.05	90.16
		Testing	92.07	92.13	91.73	91.93
	Weighted Averaging	Training	92.03	95.73	54.02	69.06
		Validation	91.67	95.12	60.0	73.58
		Testing	91.43	92.21	51.82	66.36
	ADASYN	Training	92.55	91.02	94.59	92.77
		Validation	92.69	90.41	95.31	92.79
		Testing	91.45	88.66	94.61	91.54
	Borderline SMOTE	Training	92.66	91.60	93.99	92.78
		Validation	91.96	90.48	93.99	92.20
		Testing	92.07	91.05	93.03	92.03

achieving accuracy, precision, recall and F1-Score of 92.07%, 92.13%, 91.73% and 91.93%. ResNet50+BorderlineSMOTE also attains an accuracy of 92.07%. ResNet50 obtains an accuracy of 93.21% when used with the original dataset without any re-sampling technique. However, the recall score of 68.61% is lower than ideal, suggesting that the model is missing some positive cases, i.e., chest X-rays with tuberculosis. This may be a concern in a medical setting, where identifying all positive cases is critical.

The performance of ADASYN and Borderline SMOTE, the top-performing oversampling approaches in this work, needs to be highlighted here. Borderline SMOTE targets borderline instances, boosting the model's capacity to identify complicated patterns near class boundaries. ADASYN dynamically modifies synthetic samples for sparsely inhabited areas, addressing class imbalance. These adaptive oversampling strategies highlight the necessity of tailored approaches for unbalanced datasets, contributing to improved model generalization and performance in the detection of tuberculosis.

Furthermore, upon analyzing Table 13, it is found that CNN and DenseNet 169 models have performed almost similarly in other evaluation metrics. Still, there is a wide difference in their computation time. CNN can perform comparably with DenseNet169 taking less training and testing time at 11 seconds and 3 seconds respectively whereas DenseNet169 takes 100.816 and 35.76 seconds. The figure indicates that the models have better accuracy and low loss rate. For visualization purposes, we have used Grad-CAM. While analyzing CXRs, Grad-CAM can assist in locating the precise portions of the image with the most important evidence of TB. The intensely activated regions in the heat map produced by Grad-CAM, reflect the existence of TB-related characteristics. These active areas are examined and evaluated against known TB CXR characteristics such as infiltrates, opacities, or cavities— regions that are common indications of TB infection. This facilitates the interpretation of the model's predictions and assists clinicians in making more informed decisions regarding diagnosis and treatment.

In this research work, we have performed an exhaustive and comprehensive comparative study. Three distinct comparisons are drawn in our research: one focuses on well-known pre-trained models, another centers on state-of-the-art models developed by other researchers, and the third evaluates computational time as a key metric.

4.3.3 Discussion

Comparison with DenseNet169 & ResNet50

Incorporating the notable architectures of DenseNet169 and ResNet50 in this study has been carried out to facilitate comparative analysis, as they are currently sought after in this field. Our objective is to showcase the heightened efficacy of the proposed custom CNN model by juxtaposing its results against these established benchmarks. Both these models were trained for 50 epochs. The results obtained for DenseNet169 & ResNet50 have been listed in Table 11 and 12. By leveraging domain-specific knowledge and tailored design choices, the custom CNN model performs on par with DenseNet169 in terms of accuracy, precision, recall, and F1 score, while surpassing ResNet50 in these performance metrics.

Comparison in terms of Computation Time

As can be seen in Table 13 the proposed CNN model has outperformed the pre-trained models in terms of computational efficiency. The table summarizes the computation

times, measured in seconds, for various combinations of oversampling techniques and deep learning models utilized in a research study.

Table 13: Computation time for high performing DL models.

Model	Training Time (in sec)	Testing Time (in sec)
ADASYN + CNN (Proposed Model)	**11**	**3**
Borderline SMOTE + CNN (Proposed Model)	**11.874**	**3.608**
ADASYN + DenseNet169	100.816	35.76
SMOTE + DenseNet169	112.51	31.12
W/O SMOTE + ResNet50	42.514	12.309
SMOTE + ResNet50	143.976	41.20
Borderline SMOTE + ResNet50	208.525	82.414

The proposed CNN model, when paired with ADASYN, exhibited fastest training at 11 seconds and efficient testing at 3 seconds. Similarly, when Borderline SMOTE was applied alongside the CNN model, training took 11.874 seconds, with testing requiring 3.608 seconds. In contrast, employing pre-trained models such as DenseNet169 and ResNet50 with ADASYN or SMOTE resulted in significantly longer training and testing times. Notably, the ResNet50 model without any data augmentation took 42.514 seconds for training and 12.309 seconds for testing, while combining ResNet50 with SMOTE led to further increased computation times of 143.976 seconds for training and 41.20 seconds for testing. The most time-consuming configuration was observed when using Borderline SMOTE with ResNet50, with training and testing durations of 208.525 seconds and 82.414 seconds, respectively.

5 Proposed Method 3: Deep Ensemble Model for Differential Diagnosis of COVID-19 Pneumonia

5.1 Methodology

A number of research studies have been carried out for COVID-19 detection from among a host of identical pulmonary diseases such as Pneumonia, Tuberculosis, ARDS and lung cancer. This overlap concerning the differential diagnosis for chest ailments exhibiting similar and indistinguishable symptoms is one of the major obstacles in accurately classifying suspected cases. The current study exploits different available models to further enhance their classification potential.

Our work adopts three deep learning models: Inception-V3, VGG16 and DenseNet169 that utilise CXRs to produce individual predictions unfettered by other architectures. These predictions are then put together and fed to the combined might of the ensemble model to obtain a much improved and definitive outcome. This is analogous to a situation where we give more weightage to medical diagnoses agreed upon by multiple medical specialists rather than going for a single clinical assessment. The ensemble technique can be adapted for ML and DL methods since it substantially overcomes errors in classifying target classes. Various approaches such as averaging, weighted averaging, concatenation etcetera can be put to use for achieving encouraging results. Stacking, bagging, and boosting are some other techniques that work for ML algorithms to produce augmented outcomes.

The proposed model includes the following steps: 1) Pre-processing, 2) Training individual models separately, 3) Concatenating the outcomes/feature maps of the three models

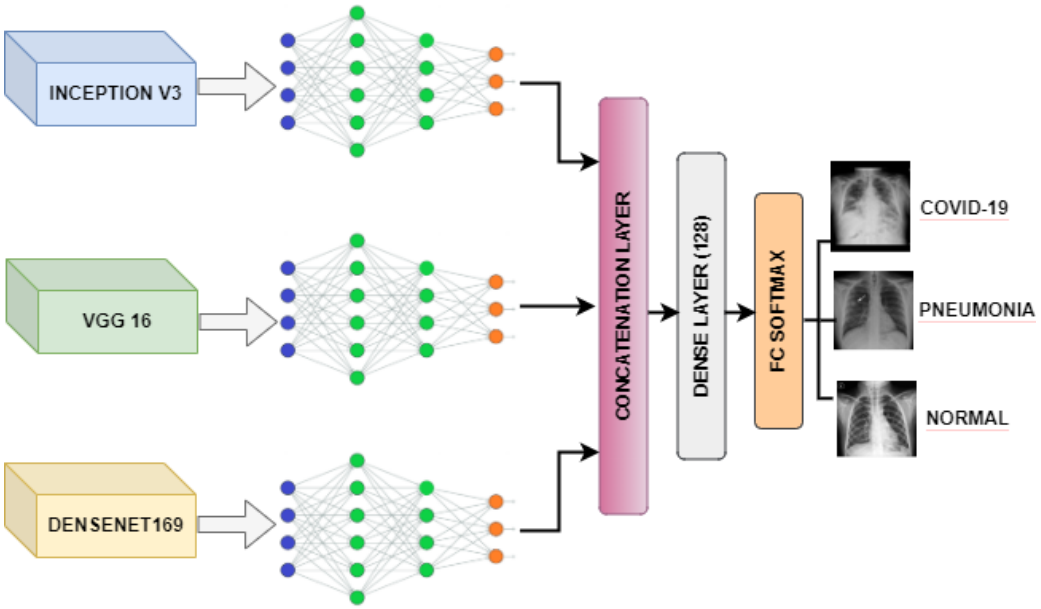

Figure 11: Block diagram of proposed Ensemble model.

to acquire an extended and comprehensive feature vector, 4) Passing the combined feature vector to ensemble model for disease classification and 5) Analysing the results obtained and carrying out comparative studies with available SOTA works. Figure 11 illustrates the Ensemble model for multi-class classification in detail. The feature maps produced by each independent pre-trained model is concatenated one after the other to generate an extensive feature vector that provides enhanced outcomes. The Ensemble model provides better accuracy since it is robust and has a comprehensive feature map since each base model extracts features and learns according to its architecture. Various base models complement each other and therefore produce better results than a single model. This feature vector is passed on to the fully connected dense layer consisting of 128 neurons with a dropout value of 0.1. The final classification is performed using Softmax activation in the last dense layer. The ensemble model has been trained for 20 epochs retaining a batch size of 64. Finally, a detection accuracy of 97.74% is achieved by the proposed model for three class classification of pulmonary diseases.

5.2 Results

Table 14: Experimental results for 3 classes of the proposed Ensemble model.

Class	Accuracy	Precision	Recall	F1-Score
Covid-19	99.48	99.47	98.95	99.21
Normal	97.91	96.04	97.98	97.00
Pneumonia	98.08	97.83	96.26	97.04

Upon analysing the pre-trained models, it is seen that DenseNet169 architecture gained the best accuracy rate of 97.39% along with precision, recall and F1-Score of 97.46%, 97.36% and 97.40% whereas Inception-V3 and VGG16 models acquired an accuracy of 95.65% and 96% respectively. As is evident from Table 15, the proposed ensemble model gave an improved outcome than any of the pre-trained models at 97.74%. The final

Table 15: Result obtained on training individual and Ensemble model

Model	Accuracy	Precision	Recall	F1-Score
Inception-V3	95.65	95.73	95.61	95.64
VGG16	96.00	96.04	96.00	96.02
DenseNet169	97.39	97.46	97.36	97.40
Ensemble	**97.74**	**97.78**	**97.73**	**97.75**

model has an accuracy rate in excess of 0.35% when compared with the best-performing individual model of DenseNet169. The results obtained for various standalone classifiers have been listed in Table 14.

5.3 Discussion

Inception-V3 model is able to successfully identify 547 CXRs and gave incorrect outcome for 28 CXRs,. Similarly, VGG16 classified 550 CXR images correctly while misdiagnosing 25 cases. The correctly classified CXRs for DenseNet169 was 560 out of 575. Our proposed Ensemble model accurately predicted 188 COVID-19 cases while misclassifying 1 each as Normal and Pneumonia. Correctly identified Normal images were 194 out of 198 while for Pneumonia 180 CXRs were correctly predicted but 7 were misclassified as Normal.

A model is said to be efficient if it obtains a high outcome for the area under ROC. It has achieved 100% for all three classes along with 100% of micro and macro averages. The Inception-V3, VGG16, DenseNet169, and Ensemble models have been trained for 20 epochs. The highest training and validation accuracy achieved was 99.57% and 98.77% whereas the loss rate for the ensemble model was 0.0181. These figures indicate that the model has better accuracy and low loss rate.

6 Proposed Method 4: An Integrated and Multitask Framework for Enhanced Segmentation and Classification of Pulmonary Diseases

6.1 Integrated Segmentation and Classification Models for Pneumonia Detection

6.1.1 Methodology

In this study, we designed a modular framework for analyzing CXR images aimed specifically at detecting various pathogens of pneumonia. The segmentation module, implemented through a custom encoder-decoder architecture with attention mechanisms, effectively delineated regions of interest, particularly the lungs, enabling a detailed exploration of anatomical variations associated with pulmonary conditions. It achieved an accuracy, IOU and Dice coefficient of 99.08%, 96.49% and 98.21% respectively. The classification module focused on analyzing global patterns in the entire image, achieving high accuracy with binary classification reaching 98.77%. The precision, recall and F1-score for the proposed model stands at 98.80%, 98.74% and 98.76%.

The research study deals with accurately identifying and classifying CXRs into target classes of a)Viral Pneumonia and Normal, b) COVID Pneumonia and Normal c) Viral Pneumonia and COVID Pneumonia. The steps involved in achieving this objective are

i) Cleaning and pre-processing the input CXR images, ii) Segmenting the ROI in input CXRs, iii) Classifying into target classes, iv) Evaluating the segmentation and classification results using various evaluation metrics such as accuracy, precision, recall, F1-score, IOU and Dice coefficient, v) Comparing the model's performance with SOTA works. The schematic representation for the proposed model is depicted in Figure 12.

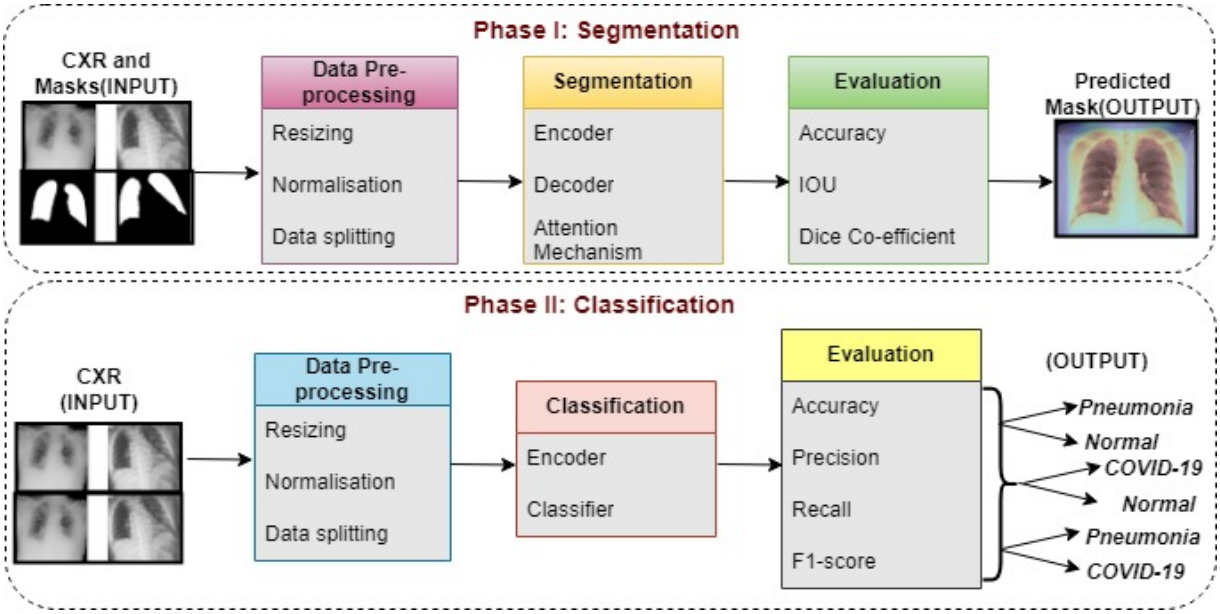

Figure 12: Architectural design for the proposed model.

6.1.2 Results

The proposed model demonstrates good performance in segmenting CXRs and classifying images, as shown in Table 16 and 17 into Normal, COVID pneumonia and Viral Pneumonia classes accurately.

Table 16: Evaluation Metrics obtained for segmentation.

Class	Accuracy (%)	IOU (%)	Dice Coefficient (%)
Pneumonia vs Normal	98.32	93.89	96.85
COVID-19 vs Normal	98.79	94.60	97.23
Pneumonia vs COVID-19	99.08	96.49	98.21

Table 17: Evaluation Metrics obtained for classification

Experiment	Accuracy	Precision	Recall	F1-score
Pneumonia vs Normal	97.36	97.34	97.30	97.32
COVID-19 vs Normal	97.16	97.38	96.91	97.11
Pneumonia vs COVID-19	98.77	98.80	98.74	98.76

The segmentation results show that the model performs excellently. The model achieved an accuracy of 98.32%, with an IOU of 93.89% and a Dice Coefficient of 96.85% for Pneumonia versus Normal, indicating strong segmentation capability. For COVID-19 versus Normal, it reached an accuracy of 98.79%, with an IOU of 94.60% and a Dice Coefficient of 97.23%, reflecting superior segmentation performance. The classification between Pneumonia and COVID-19 achieved an accuracy of 99.08%, an IoU of 96.49%, and a Dice coefficient of 98.21%. Overall, the model effectively distinguishes disease areas from normal tissue with high accuracy and consistency.

In the classification experiments, the model demonstrated strong performance across all tasks. For distinguishing Pneumonia from Normal cases, it achieved an accuracy of 97.36%, with a precision of 97.34%, recall of 97.30%, and an F1-score of 97.32%. This indicates that the model reliably identifies Pneumonia while maintaining a balance between precision and recall. For COVID-19 versus Normal classification, the model reached an accuracy of 97.16%, with a precision of 97.38%, recall of 96.91%, and an F1-score of 97.11%. These metrics reflect the model's effectiveness in correctly identifying COVID-19 cases with minimal false positives and negatives. In the COVID-19 versus Pneumonia classification, the model achieved an impressive accuracy of 98.77%. This high accuracy indicates that 98.77% of the images were correctly classified into either the COVID-19 or Pneumonia categories. The precision of 98.80% means that when the model predicted an image as either COVID-19 or Pneumonia, it was correct 98.80% of the time. This high precision suggests that the model effectively minimizes false positives, ensuring that misclassifications are rare. The recall of 98.74% demonstrates that the model successfully identified 98.74% of the actual COVID-19 or Pneumonia cases among all true positives. This high recall indicates the model's strong sensitivity in detecting both diseases, leaving few cases undetected. The F1-score of 98.76% provides a balanced measure of precision and recall, reflecting the model's overall excellence in distinguishing between COVID-19 and Pneumonia. A high F1-score indicates that the model maintains a good trade-off between precision and recall, ensuring both accurate and comprehensive identification of the diseases. The confusion matrices provide a visual breakdown of the correctly classified classes, along with the false positives and false negatives.

6.1.3 Discussion

In this study, we implemented separate modules for segmentation and classification of CXR images to allow for independent optimization of each task. The segmentation module, although not directly influencing the classification outcomes, played a crucial role in identifying and delineating regions of interest, such as lung areas. This enabled us to examine anatomical variations linked to different pulmonary conditions. Conversely, the classification module concentrated on the overall patterns present in the entire image to categorize cases into COVID-19, pneumonia, or normal classes. This modular design not only allowed for specialized optimization but also offered a flexible framework, making it adaptable for future advancements in segmentation and classification techniques. While the segmentation results were not utilized in the classification process, they add a complementary layer of analysis, potentially enhancing clinical interpretation and guiding further research.

6.2 PulmoMTL-Net: A Multitask Learning Framework for Enhanced Segmentation and Classification of Pulmonary Diseases

6.2.1 Methodology

In the proposed methodology shown in Figure 13, a multi-task learning model is designed to handle three tasks: image reconstruction, segmentation, and classification. The encoder extracts hierarchical features from the input image, followed by two decoders—one for reconstructing the original image and the other for generating multi-class segmentation masks. Additionally, a classification head processes the segmentation output, employing DenseNet169 for accurate classification of pulmonary diseases. The model integrates Gaussian blur to enhance segmentation performance, improving the overall classification process.

Architecture

Multi-Task Learning is a ML paradigm in which multiple learning tasks are solved simultaneously while sharing representations. By leveraging commonalities and differences across tasks, MTL aims to improve generalization performance on individual tasks compared to learning them individually[44].

In Multi-Task Learning (MTL), let $\mathcal{T} = \{T_1, T_2, \ldots, T_n\}$ represent a set of n tasks. Each task T_i has its own dataset $D_i = \{(x_j^i, y_j^i)\}_{j=1}^{m_i}$, where x_j^i is the input feature vector and y_j^i is the target output. The total loss L for MTL can be defined as the weighted sum of the individual task losses:

$$L(\theta) = \sum_{i=1}^{n} \lambda_i L_i(\theta, D_i)$$

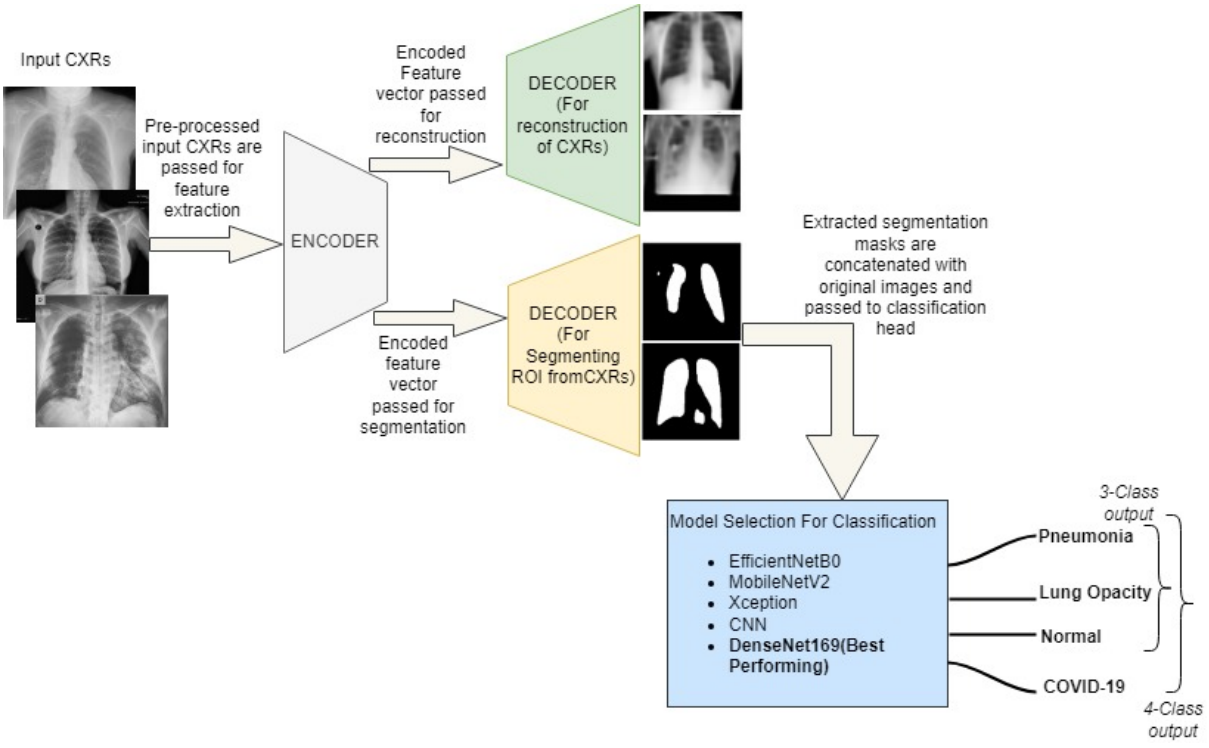

Figure 13: Block diagram of proposed PulmoMTL-Net model.

where θ represents the model parameters, L_i is the loss function for task T_i, and λ_i is the weight assigned to task T_i. In MTL, the model typically consists of a shared representation $f(x;\theta)$ that extracts common features for all tasks. The output for each task can be expressed as:

$$\hat{y}_j^i = g_i(f(x_j^i;\theta);\theta_i)$$

where g_i is a task-specific head for task T_i that processes the shared representation.

MTL aims to optimize the shared parameters θ across all tasks, with the optimization process expressed as:

$$\theta^* = \arg\min_\theta L(\theta) = \arg\min_\theta \sum_{i=1}^{n} \lambda_i L_i(\theta, D_i)$$

Using gradient descent, the update rule for parameters can be formulated as:

$$\theta \leftarrow \theta - \eta \nabla L(\theta)$$

where η is the learning rate and $\nabla L(\theta)$ is the gradient of the total loss for the parameters.

In the proposed study, the model includes segmenting the lung region from CXRs as the segmentation task and classifying the type of infection from the images as the classification task. Additionally, the CXRs are also reconstructed from the encoded feature maps. Incorporating a reconstruction task in a multi-task learning model enhances feature learning by extracting essential details from input data, which improves segmentation and classi- fication performance. It acts as a regularization mechanism, reducing over- fitting, especially in small datasets.The reconstruction also provides contextual insights and visual feedback, aiding interpretability and trust in clinical applications. Overall, this approach results in higher accuracy across tasks by effectively balancing multiple objectives. In this regard, the shared representation can extract features from the images that are useful for recon- struction, segmenting lung structures and classifying lung/pulmonary dis- ease, ultimately improving overall performance.

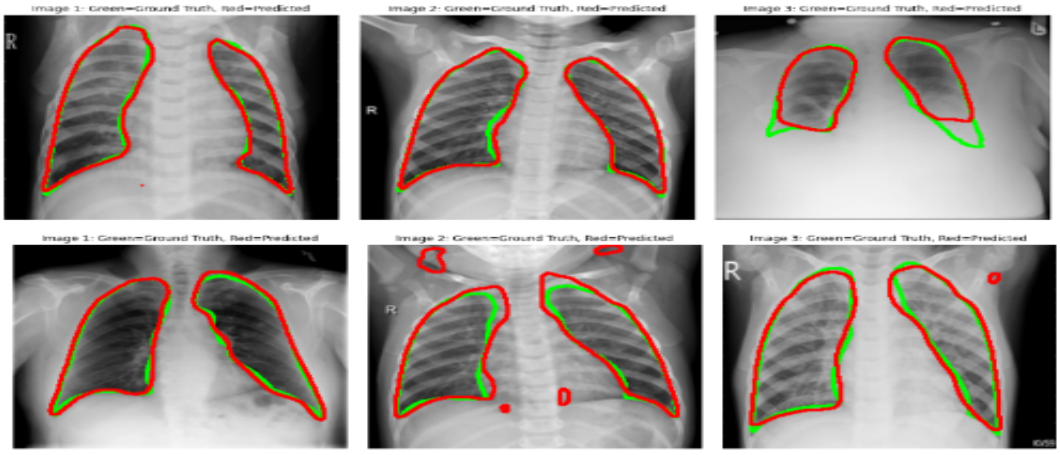

Figure 14: Input CXRs with contour lines. i) Red line shows predicted ROI and ii) Green lines show the ground-truth

6.2.2 Results

The results from the various models highlight their performance across reconstruction, segmentation, and classification tasks. In classification, the proposed model achieves training accuracy of 99.36% for the three-class problem and 99.23% for the four-class problem, outperforming EfficientNetB0, MobileNetV2, and Xception, which have slightly lower accuracies as shown in Table18. The training classification loss for the proposed model is 0.0272 for three classes and 0.0268 for four classes, indicating a balance between high accuracy and low loss. The test results shown in Table 21 demonstrate that PulmoMTL-Net with DenseNet169 as the classifier is the best performing model attaining test accuracy, precision, recall, F1-score of 96.55%, 96.59%, 96.55%, 96.56% for three class and 95.38%, 95.42%, 95.38% and 95.37% respectively for four class problems. In segmentation, the proposed PulmoMTL-Net model exhibits test accuracy of 94.51% for the three-class problem and 95.66% for the four-class problem. The Intersection over Union (IoU) improves from 85.11% for three classes to 90.17% for four classes, showing better overlap between predicted and actual masks. The Dice coefficient further reinforces the model's efficacy, reaching 91.94% for three classes and 94.81% for four classes. For reconstruction, the proposed model demonstrates effective performance with a training MSE of 0.0013 for three classes and 0.0009 for four classes, validation MSE at 0.0058 and 0.0008 and testing MSE at 0.0042 and 0.0007 respectively. The results are listed in Table 19. Overall, the proposed PulmoMTL-Net model excels in multi-task learning, achieving high accuracy and low reconstruction error, positioning it as a robust solution for medical image analysis in pulmonary disease diagnosis.

The original CXRs serve as a baseline, representing the model's input data. The reconstructed images produced by the model's decoder are intended to simulate the original input by retaining essential features while improving finer details. The visual comparison of the original and reconstructed images demonstrates the model's capacity to preserve essential information during the reconstruction process, ensuring that the increased features contribute to the following segmentation and classification tasks. Figure 14 presents a visual comparison of the model's performance in segmenting the ROIs in CXR images. The red contour lines delineate the areas identified by the model as significant, highlighting the segmented lung regions or other key areas relevant for classification. Green contour lines, on the other hand, represent the ground truth, showing the manually annotated regions used as the standard reference. The overlapping areas between the red and green contours indicate the model's accuracy in correctly identifying these regions, while any discrepancies highlight areas where the model's predictions differ from the actual ground truth.

6.2.3 Discussion

In conclusion, the proposed PulmoMTL-Net model demonstrates promising performance in the multitask learning framework for pulmonary disease classification and segmentation. The model achieved an impressive three-class classification accuracy of 96.55%, precision of 96.59%, recall of 96.55%, and F1 score of 96.56%, alongside a four-class classification accuracy of 95.38%, precision of 95.42%, recall of 95.38%, and F1 score of 95.37%. For segmentation tasks, the model yielded an Intersection over Union (IoU) of 85.11% and a Dice coefficient of 91.94% as shown in Table 20. Additionally, the reconstruction Mean Squared Error (MSE) was recorded at 0.0042 for the three-class scenario and 0.0007 for the four-class scenario.

Table 18: The following table summarizes the key performance indicators, including classification and segmentation accuracy and loss, for both training and validation datasets across various experiments. Apart from Encoder-Decoder architecture, different classifiers have been utilized and the best performing classification architecture has been selected for the proposed PulmoMTL-Net model.

Method	**Classification Accuracy**	**Classification Loss**	**Segmentation Accuracy**	**Segmentation Loss**
EfficientNetB0	99.28	0.0215	95.39	0.0958
	30.18	6.1165	95.38	0.9781
MobileNetV2	99.33	0.0190	95.95	0.0818
	90.32	0.6722	95.93	0.0814
Xception	99.45	0.0192	94.59	0.1185
	93.77	0.3824	95.12	0.1058
CNN	96.00	0.1069	95.74	0.0909
	89.29	0.3623	95.35	0.1006
Proposed [**Three Class**]	99.36	0.0272	95.43	0.0909
	94.93	0.3440	94.50	0.1210
[**Four Class**]	99.23	0.0268	96.48	0.0407
	92.29	0.2844	95.66	0.0428

Table 19: MSE values for different models across training, validation, and test phases. The table includes test MSE for both three-class and four-class classification tasks for the proposed model.

Model	**Training MSE**	**Validation MSE**	**Test MSE**	
EfficientNetB0	0.0018	0.0017	0.0017	-
MobileNet	0.0015	0.0012	0.0011	-
Xception	0.0024	0.0013	0.0018-	-
Custom NN	0.0014	0.0018	0.0019	-
Proposed(Three Class)	0.0013	0.0058	0.0042	-
Proposed(Four Class)	0.0009	0.0008	-	0.0007

Table 20: Segmentation Results for the proposed PulmoMTL-Net model. The table shows the evaluation metrics for both; three-class as well as four-class problem.

Model	Loss	Accuracy	IoU	Dice Coefficient
3-Class	0.1210	94.51	85.11	91.94
4-Class	0.0428	95.66	90.17	94.81

Table 21: Classification results obtained for various experimental models along with the proposed model.

Model	No. of Classes	Accuracy	Precision	Recall	F1-Score
EfficientNetB0	3	85.06	86.13	85.06	85.01
MobileNetV2	3	91.49	91.49	91.49	91.49
Xception	3	93.79	93.81	93.79	93.77
CNN	3	91.00	92.00	92.00	92.00
Proposed	3	**96.55**	**96.59**	**96.55**	**96.56**
Model	4	**95.38**	**95.42**	**95.38**	**95.37**

The effectiveness of PulmoMTL-Net highlights its potential for real-time applications in medical diagnostics, particularly in enhancing the accuracy of disease detection and facilitating timely interventions. However, the model's reliance on high-quality annotated datasets may limit its generalizability in diverse clinical settings. Future research should focus on exploring the integration of larger and more varied datasets to improve the model's robustness and accuracy further.

7 Proposed Method 5: Performance Analysis of Deep Neural Networks and Machine Learning Algorithms for Pulmonary Disease Classification

7.1 Methodology

This research work proposes a hybrid model as shown in Figure 15, that utilizes a Convolutional Neural Network (CNN) for feature extraction and a machine learning classifier for the detection of COVID-19 from CXRs. The CNN is responsible for extracting relevant features from the X-ray scans of patients, while the machine learning classifier is trained to differentiate between COVID-19, pneumonia, and normal cases. The study consists of the following steps: i) Pre-processing, ii) Feature Extraction using various CNN models such as DenseNet121, DenseNet169, ResNet50, Xception and such. iii) Classification of extracted features into appropriate classes with the help of ML classifiers: RF, SVM and LR. Encouraging results were achieved which have been listed in section 7.2, 7.3

Pre-processing

The images in the datasets were resized to a size of 224 X 224 and were then normalized i.e. the pixel values were converted to lie within the range of [0,1]. Various approaches for augmenting data have been used in this work. Images were rotated clockwise with an angle of 5 degrees, scaled in the range of 0.9 to 1.2, flipped horizontally and translated

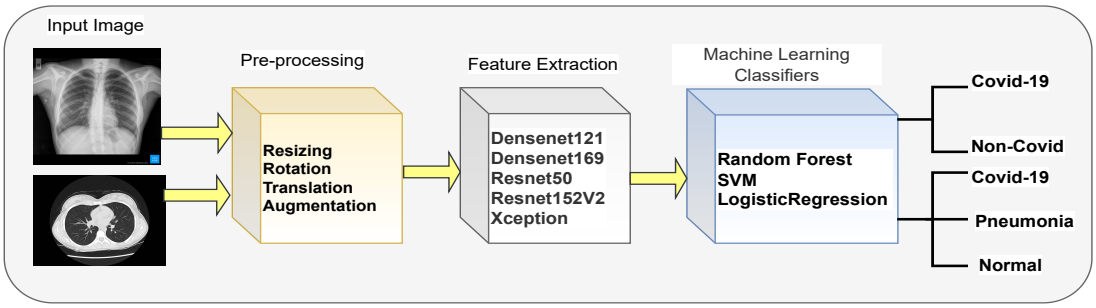

Figure 15: Architecture of CNN based feature extractor + ML Classifier Model.

horizontally as well as vertically by -5 to 5%. For training and testing the model the datasets were randomly divided into two parts in a ratio of 80% and 20%.

Feature Extraction

Transfer learning approach has been employed in this work along with different CNN models already trained on Imagenet dataset: DenseNet121, DenseNet169, ResNet50, ResNet152v2 and Xception. In DenseNet architecture each layer has a direct connection to all the remaining layers in the network. If the architecture has 'm' number of layers then there are m(m+1)/2 connecting lines between the layers. The layers contained within a block allow others to access and use the features extracted by them. Therefore, DenseNet needs less number of parameters leading to a compact performance that produces better results as compared to other standard architectures. DenseNet121, having [6,12,24,16] layers in the four dense blocks as well as DenseNet-169 with [6, 12, 32, 32] layers have been used in this work.

Residual network or ResNet was developed by Ren et al. [18] to overcome the problems that arose in training deep networks. The 34 layers model, inspired by VGG-19, consists of skip connections which help in solving the vanishing gradient problem. Two versions of ResNet: ResNet152v2 & ResNet 50 have been used as feature-extractor here.

According to Chollet[10] the Xception architecture, which has been based on the Inception CNN model, is a continuous stack of residually connected depth-wise separable convolution layers. Because of this, the architecture is exceedingly simple to define and change. This model gives a better accuracy on Imagenet dataset than InceptionV3 because of better parameter usage.

Classification

The classification techniques entail foretelling a particular result based on a given set of inputs. The technique employs a training set comprising of a collection of features to forecast the final result. We have used conventional ML methods such as SVM, LR and RF to accurately group the extracted features into appropriate classes. From the datasets being used, 80% of the images were taken for training whereas the remaining 20% were employed for testing purposes. All three classifiers were used in combination with a feature extractor model. The model was executed using five different datasets and a total of 5*3*6 experiments were carried out for COVID-19 identification.

7.2 Results

Binary as well as multi-class classification were carried out in this research paper by using different combinations of feature extraction mechanisms and ML classifiers. In total 5 X 3 X 6=90 experiments were carried out and the model was executed on five datasets. The final outputs obtained thereafter have been listed in Table 22, 23, 24, 25, 26, 27.

7.3 Discussion

SARS-Cov-2 CT dataset[DS2], Covid-19 and Common Pneumonia dataset[DS1] as well as Chest Xray-Pa dataset [DS4] have been taken up for detecting COVID-19 from Normal / Pneumonia / Non-covid cases. We have partially utilized DS4, since we require only two classes for binary classification although it contains three target classes: Covid-19, Normal & Pneumonia.

Table 22: Binary Classification using DS2: Results Achieved

DATASET	FEATURE EXTRACTOR	CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
SARS-Cov-2 Dataset[DS2]	DenseNet121	RF	91.95	91.98	91.94	91.95
		SVM	98.39	98.39	98.39	98.39
		LR	98.39	98.39	98.39	98.39
	DenseNet169	RF	91.35	91.35	91.35	91.35
		SVM	98.39	98.39	98.38	98.39
		LR	97.98	98.00	97.97	97.98
	ResNet50	RF	88.12	88.12	88.13	88.12
		SVM	95.97	95.97	95.97	95.97
		LR	93.15	93.16	93.15	93.15
	ResNet152V2	RF	89.33	89.34	89.32	89.33
		SVM	95.57	95.60	95.56	95.57
		LR	95.17	95.20	95.15	95.16
	Xception	RF	89.33	89.34	89.34	89.33
		SVM	95.37	95.41	95.35	95.36
		LR	94.96	95.04	94.95	94.96

Table 23: Binary Classification using DS1: Results Achieved

DATASET	FEATURE EXTRACTOR	CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
Covid-19 & Common Pneumonia dataset[DS1]	DenseNet121	RF	100	100	100	100
		SVM	100	100	100	100
		LR	100	100	100	100
	DenseNet169	RF	100	100	100	100
		SVM	100	100	100	100
		LR	100	100	100	100
	ResNet50	RF	100	100	100	100
		SVM	100	100	100	100
		LR	100	100	100	100
	ResNet152V2	RF	99.05	99.29	98.59	98.92
		SVM	100	100	100	100
		LR	100	100	100	100
	Xception	RF	99.52	99.64	99.30	99.47
		SVM	100	100	100	100
		LR	100	100	100	100

Table 22,23,24 presents the accuracy, precision, recall and F1-score for all combinations of feature extractor and classifier. For DS2 the best accuracy result has been achieved for DenseNet121 and SVM, DenseNet121 and LR & DenseNet169 and SVM model - 98.39% followed by DenseNet169 and LR which obtained 97.98%. The confusion matrix clearly presents the expected as well as predicted values of the model for DS2. In DS1

Table 24: Binary Classification using DS4: Results Achieved

DATASET	FEATURE EXTRACTOR	CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
Chest-xray Pa Dataset[DS4]	DenseNet121	RF	95.73	95.72	95.74	95.73
		SVM	98.52	98.50	98.55	98.52
		LR	98.52	98.51	98.52	98.52
	DenseNet169	RF	96.06	96.04	96.11	96.06
		SVM	98.03	98.00	98.07	98.03
		LR	98.36	98.33	98.38	98.35
	ResNet152V2	RF	93.77	93.74	93.81	93.76
		SVM	97.70	97.68	97.72	97.70
		LR	97.54	97.51	97.57	97.53
	ResNet50	RF	91.14	91.19	91.25	91.14
		SVM	95.08	95.05	95.10	95.07
		LR	94.75	94.73	94.75	94.74
	Xception	RF	90.98	91.06	91.11	90.98
		SVM	97.37	97.35	97.40	97.37
		LR	97.37	97.36	97.38	97.37

DenseNet121, DenseNet169 and ResNet50 together with ML classifiers RF, SVM and LR gave an excellent accuracy of 100% each. Other pairings such as ResNet152V2 along with SVM and LR and Xception with SVM and LR also reached 100%.

DS4 achieved best result for DenseNet121 and SVM & DenseNet169 with SVM & LR pairing as can be seen in the Table 24. The fusion of other feature extraction methods such as Xception, ResNet50 and ResNet 152V2 fails to keep up with the accuracy threshold attained by DenseNet121 and DenseNet169 in the remaining experiments which were conducted. Although DS5 comprises of three categories of images, we made use of a subset of the classes and trained and tested the model for binary classification. Accuracy score of 99.80% was achieved while distinguishing between COVID-19 and normal groups.

Three datasets, Mendeley's Covid Pneumonia Dataset [DS3], Covid-Pneumonia-Normal-chest-x-raypa Dataset [DS4] and Covid-19 Radiography Dataset [DS5] were put to use in this research work. All these datasets have three target classes each of Normal, Pneumonia and Covid-19 image types. While working with DS5, the highest detection rate was that of DenseNet121 and SVM, DenseNet121 and LR and DenseNet169 and SVM fusion models: 98.82%, 98.56% and 98.56% respectively. The deep learning models could detect COVID-19 cases with 100% accuracy for DenseNet121 with SVM, DenseNet121 with LR and DenseNet169 with SVM in DS3. DS5 closely followed with 98.82% and 98.56% for DenseNet121 and SVM and DenseNet169 and SVM respectively. DS4 did not fare very well as compared to the above-mentioned results of other datasets. The highest accuracy achieved for the DS4 is 95.51% which was obtained by the combination of DenseNet121 and LR. This was followed by the pair of DenseNet121 and SVM which produced an accuracy score of 95.08%.

As evident from Table 22,23,24 & Table 25,26,27, we can conclude that out of the different combinations of deep learning models employed in this paper, DenseNet121 and DenseNet169, when applied with SVM and LR give promising results consistently across divergent datasets.

Table 25: Results achieved for Multi-class Classification using DS5

DATASET	FEATURE EXTRACTOR	CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
Covid-19 Radiography Dataset[DS5]	DenseNet121	RF	96.47	96.58	96.51	96.52
		SVM	98.82	98.86	98.83	98.84
		LR	98.56	98.60	98.57	98.58
	DenseNet169	RF	97.12	97.24	97.16	97.18
		SVM	98.56	98.60	98.58	98.59
		LR	98.43	98.47	98.45	98.46
	ResNet50	RF	91.64	91.91	91.68	91.74
		SVM	96.74	96.75	96.73	96.74
		LR	95.69	95.75	95.73	95.73
	ResNet152V2	RF	95.82	95.95	95.88	95.92
		SVM	97.52	97.61	97.55	97.58
		LR	97.52	97.61	97.56	97.58
	Xception	RF	94.13	94.27	94.17	94.21
		SVM	96.61	96.66	96.67	96.66
		LR	95.69	95.81	95.74	95.76

Table 26: Results achieved for Multi-class Classification for DS3

DATASET	FEATURE EXTRACTOR	CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
Mendeley's CT Scan Dataset [DS3]	DenseNet121	RF	99.83	99.84	99.84	99.84
		SVM	100	100	100	100
		LR	100	100	100	100
	DenseNet169	RF	99.91	99.91	99.90	99.91
		SVM	100	100	100	100
		LR	99.91	99.92	99.91	99.92
	ResNet50	RF	99.74	99.76	99.74	99.75
		SVM	100	100	100	100
		LR	98.55	98.65	98.63	98.64
	ResNet152V2	RF	98.72	98.79	98.66	98.72
		SVM	99.83	99.84	99.82	99.83
		LR	99.83	99.84	99.82	99.83
	Xception	RF	94.12	94.26	94.17	94.21
		SVM	96.60	96.66	96.66	96.65
		LR	95.69	95.80	95.73	95.76

Table 27: Results achieved for Multi-class Classification for DS4

DATASET	FEATURE EXTRACTOR	CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
Chest-Xray-Pa Dataset[DS4]	DenseNet121	RF	92.45	92.71	92.49	92.45
		SVM	95.08	95.07	95.08	95.04
		LR	95.51	95.53	95.52	95.48
	DenseNet169	RF	92.34	92.76	92.40	92.36
		SVM	94.53	94.51	94.50	94.48
		LR	94.64	94.62	94.62	94.60
	ResNet50	RF	88.08	88.49	88.15	88.12
		SVM	89.39	89.35	89.36	89.35
		LR	87.86	87.35	87.96	87.91
	ResNet152V2	RF	90.27	90.80	90.36	90.32
		SVM	94.31	94.36	94.31	94.29
		LR	94.75	94.82	94.76	94.72
	Xception	RF	90.60	91.18	90.66	90.65
		SVM	92.34	92.41	92.35	92.31
		LR	91.91	91.98	91.90	91.88

8 Conclusions and Future Works

8.1 Conclusions

The presented research explores the potential of advanced DL and ML techniques for the classification and segmentation of pulmonary diseases, emphasizing their critical role in

medical diagnostics. Pulmonary diseases such as Pneumonia, TB and Lung Cancer have suddenly come into the spotlight with the emergence of COVID-19. Earlier pneumonia was considered a silent killer, affecting millions of children and adults annually. Healthcare systems all over the world were tested and overwhelmed during the COVID-19 pandemic, as overlapping symptoms with other lung diseases, such as pneumonia, tuberculosis, and viral respiratory infections, further exacerbated the situation. This overlap often led to delays in accurate diagnosis, misclassification of diseases, and strain on diagnostic resources, complicating patient management and treatment outcomes. According to the WHO, the number of people affected by pulmonary diseases is substantial and is expected to increase by almost 10% every year. Therefore a CAD for lung diseases is of utmost importance for assisting medical personnel in the accurate and timely detection of ailments.

As an initial step of the research, handcrafted features extracted using Gabor filters, LBP, HOG and GLCM were successfully integrated with DenseNet121 features in the proposed model to significantly enhance the classification performance. By combining the detailed texture and structural information from handcrafted features with the advanced DL capabilities of DenseNet121, this approach improved both feature extraction and representation. When used in conjunction with traditional ML algorithms like SVM, KNN, DT, NB, RF and MLP, this hybrid method achieved superior classification accuracy and robustness. The integration of these diverse feature sets provided a more thorough analysis of the data, resulting in more precise and reliable lung disease classification compared to standalone models.

The next proposed work employed a CNN-based feature-extractor with a ML classifier model for detecting COVID-19 from normal/Pneumonia scans. The model also successfully classified COVID-19, Pneumonia and normal images. Five feature extraction methods, including DenseNet121, DenseNet169, ResNet50, ResNet152V2 and Xception, were utilized along with RF, SVM and LR classifiers. The model was tested on five datasets to analyze its generalization capability. Finally, it was concluded that DenseNet121 and DenseNet169 methods give the best results when employed along with SVM and LR classifiers, followed by the ResNet + SVM fusion model. An accuracy of 100% was obtained for DS1 by DenseNet121, DenseNet169 and ResNet50 when teamed up with ML classifiers RF, SVM and LR considering binary classification. The accuracy achieved for DS3 and DS5 datasets (having three classes) was 100%, 98.82% and 98.56%. Looking at the model's performance, it could be safely concluded that the investigated model can detect COVID-19 from CXR and CT scan images with less error.

In the next research study, three Deep CNN models - Inception V3, VGG16 and DenseNet169 were successfully ensembled by using the concatenation method for the classification of Covid-19 Pneumonia, Viral Pneumonia and healthy cases. The model achieved an accuracy, precision, recall and F1-Score of 97.74%, 97.78%, 97.73% and 97.75%. It could be concluded that the model has high performance and could be used successfully for detecting and distinguishing between COVID-19 and Viral Pneumonia.

In the next stage of research, we proposed the CNN-LSTM model for enhancing the classification of multi-class pulmonary diseases. Alongside, we utilized two other baseline models, CNN and CNN-Bi LSTM, for comparative analysis. The hybrid CNN-LSTM model emerged as the most effective, demonstrating robust performance in disease classification with accuracy, precision, recall, and F1 Score of 96.24%, 96.44%, 96.62%, and 96.40%, respectively. The experimental evaluation conducted showcased improved accuracy compared to existing state-of-the-art models, validating the proposed model's capability to achieve well-balanced results in terms of AUC-ROC and F1 scores.

We further integrated CNN and GRU combining the strengths of both architectures. The CNN excels at feature extraction from raw pixel data, while the GRU specializes in sequential analysis and capturing temporal dependencies. By leveraging the feature representations learned by the CNN and further refining them through sequential processing by the GRU, the developed model was able to achieve enhanced discriminative power and classification performance. Data augmentation through transformations such as rotation, scaling, and flipping further improved the model's robustness and generalization ability by exposing it to a wider variety of image variations. Moreover, contrast stretching and histogram equalization, enhanced the quality of input images, aiding in the extraction of relevant features and improving the model's performance. In real-world settings, the model could significantly improve TB detection, particularly in resource-limited areas, by aiding early diagnosis and treatment, ultimately impacting global TB control efforts. In communities with a high TB incidence and limited access to qualified radiologists, this model might be incorporated into mobile health units or telemedicine systems, allowing for remote TB screening and diagnosis. This would allow for more timely intervention, especially in underserved communities, potentially lowering TB death and morbidity rates. Furthermore, the LIME package's ability to highlight significant features utilized in predictions provides the model with transparency and interpretability, both of which are critical for garnering the trust of healthcare providers and patients. Overall, our study contributes to the ongoing efforts to develop sophisticated AI-driven solutions for TB detection, with the ultimate goal of enhancing healthcare outcomes and addressing public health challenges.

Healthcare industry including, drug therapy for TB was significantly impacted due to Covid-19. Pandemic-related restrictions resulted in delays in diagnosing and initiating medical care for TB-infected patients, leading to a significant increase in mortality rates. Reports indicated reduced TB testing in countries with historically high TB cases. The Global Tuberculosis Report 2021 revealed a 41% decline in TB case notifications in India out of a total of 1.3 million cases. This research aims to enhance the detection rate of TB-infected patients by leveraging CXRs. The next proposed method investigated the efficacy of a custom CNN model in identifying TB from patients' CXRs. The study also compared the performance of the proposed model with two existing pre-trained CNNs, namely DenseNet169 and ResNet50, to determine its suitability for the task. The dataset used in the study was imbalanced therefore, various oversampling techniques, including SMOTE, ADASYN, borderline SMOTE, and weighted averaging, were employed to address the issue. Furthermore, the study employed Grad-CAM to depict the regions of interest, allowing for a visual representation of the specific areas in the analyzed data that contributed most significantly to the model's predictions or outcomes. Based on the results, the custom CNN model performed best in terms of computation time, while the accuracy was similar for both DenseNet169 and custom CNN. The study also revealed that oversampling techniques significantly improved the performance of the models. Overall, the research underscores the potential of custom CNN model in TB detection from CXRs, offering a promising avenue to enhance disease diagnosis, particularly in resource-limited settings.

In the next study, we implemented separate modules for segmentation and classification of CXRs to allow for independent optimization of each task. The segmentation module, although not directly influencing the classification outcomes, played a crucial role in identifying and delineating regions of interest, such as lung areas. This enabled us to examine anatomical variations linked to different pulmonary conditions. Conversely,

the classification module concentrated on the overall patterns present in the entire image to categorize cases into COVID-19, pneumonia, or normal classes. This modular design not only allowed for specialized optimization but also offered a flexible framework, making it adaptable for future advancements in segmentation and classification techniques. While the segmentation results were not utilized in the classification process, they added a complementary layer of analysis, potentially enhancing clinical interpretation and guiding further research.

The final proposed PulmoMTL-Net model demonstrated promising performance in the multitask learning framework for pulmonary disease segmentation and classification. The model achieved an impressive three-class classification accuracy of 96.55%, precision of 96.59%, recall of 96.55% and F1 score of 96.56%, alongside a four-class classification accuracy of 95.38%, precision of 95.42%, recall of 95.38%, and F1 score of 95.37%. For segmentation tasks, the model yielded an Intersection over Union (IoU) of 85.11% and a Dice coefficient of 91.94%. Additionally, the reconstruction Mean Squared Error (MSE) was recorded at 0.0042 for the three-class scenario and 0.0007 for the four-class scenario. The effectiveness of PulmoMTL-Net highlights its potential for real-time applications in medical diagnostics, particularly in enhancing the accuracy of disease detection and facilitating timely interventions.

8.2 Future Works

Future research will focus on incorporating additional disease classes into the classification models to enhance their capability to handle more diverse pulmonary diseases. Expanding the dataset size and diversity will be a critical step toward improving the robustness and generalizability of the models across varied clinical settings.

The proposed models can be further optimized by experimenting with advanced architectures and feature extraction methods. Integrating multi-modal data, such as combining CXRs with patient metadata or CT scans, may provide additional insights to improve classification performance. Addressing challenges like class imbalance through advanced sampling techniques and weighted loss functions will also be a key area of exploration. To enhance segmentation quality, future work could investigate the use of more sophisticated segmentation models, refined post-processing techniques, and attention mechanisms. The segmentation and classification modules may also be integrated more closely, enabling segmentation results to directly influence classification accuracy.

For real-time applications, exploring lightweight models and hardware acceleration techniques can reduce computation time, making the solutions more accessible in resource-constrained environments. The models could also be deployed in mobile health systems or telemedicine platforms for remote diagnostics, especially in underserved communities. Improving model interpretability remains an essential goal. Expanding the use of tools like Grad-CAM and LIME to highlight regions critical to model predictions will provide better insights for healthcare professionals, ensuring the models' predictions are trusted and actionable in clinical practice. Testing the models with different optimizers and hyperparameter tuning strategies may provide insights into further improving their efficiency and performance. Future work will also explore multi-task learning frameworks that extend beyond segmentation and classification to include prediction of disease progression or treatment outcomes. By addressing these avenues, future research aims to develop more robust, adaptable and impactful AI-driven solutions for pulmonary disease detection and diagnosis.

9 Publications

The following works have been accepted and presented in international conferences:

- Shekar, B.H., Mannan, S., Hailu, H. (2023). **Pneumonia Detection Using Deep Learning Based Feature Extraction and Machine Learning**". In: Gupta, D., Bhurchandi, K., Murala, S., Raman, B., Kumar, S. (eds) Computer Vision and Image Processing. CVIP 2022. Communications in Computer and Information Science, vol 1777. Springer, Cham. https://doi.org/10.1007/978-3-031-31417-9_45
- B. H. Shekar, Shazia Mannan, Habtu Hailu, C. Krishna Mohan, and C. Linga Reddy "**Performance analysis of deep neural networks for COVID-19 detection from chest radiographs**", Proc. SPIE 12701, Fifteenth International Conference on Machine Vision (ICMV 2022), 127010R (7 June 2023); https://doi.org/10.1117/12.2679620
- B. H. Shekar and S. Mannan, "**A Deep Ensemble Model for Differential Diagnosis of Covid-19 Pneumonia using Thoracic Radiography**," 2023 IEEE 10th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2023, pp. 590-596, doi: 10.1109/SPIN57001.2023.10116517.
- S. Mannan and B. H. Shekar, "**Deep Learning Fusion: CNN-GRU Model for Tuberculosis Detection Enhanced with LIME Interpretability**," 2024 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI), Prayagraj, India, 2024, pp. 1-8, doi: 10.1109/CVMI61877.2024.10782469.
- B. H. Shekar and S. Mannan, " **Integrating Handcrafted Textural and Deep Learning Features for Improved Multi-Class Lung Disease Classification in Computed Tomography Imaging**". Presented at 3rd International Conference on Paradigm Shifts in Communication, Embedded Systems, Machine Learning and Signal Processing (PCEMS 2024), VNIT Nagpur, Maharashtra - 440010, INDIA
- Shazia Mannan and B. H. Shekar, "**Differential Diagnosis of Pulmonary Diseases using Convolutional Neural Network with LSTM Architecture**". Presented at the 3rd Congress on Smart Computing Technologies (CSCT 2024), National Institute of Technology, Sikkim.
- B. H. Shekar and S. Mannan, "**Improving Pneumonia Diagnosis with Integrated Segmentation and Classification Models**". Presented at the International Conference on Responsible Artificial Intelligence (ICRAI) 2024, organized by Mangalore University, Karnataka.

The following works have been communicated to journals:

- B. H. Shekar and S. Mannan, "**Tuberculosis Detection with Customized CNN and Oversampling Techniques: A Deep Learning Approach**". Submitted to Discover Artificial Intelligence. Under Review.
- B. H. Shekar and S. Mannan, "**PulmoMTL-Net: A Multitask Learning Framework for Enhanced Segmentation and Classification of Pulmonary Diseases**". Submitted to the International Journal of Computers and Applications. Under Review.

References

[1] Rahib H Abiyev and Abdullahi Ismail. Covid-19 and pneumonia diagnosis in x-ray images using convolutional neural networks. *Mathematical Problems in Engineering*, 2021, 2021.

[2] Priya Aggarwal, Narendra Kumar Mishra, Binish Fatimah, Pushpendra Singh, Anubha Gupta, and Shiv Dutt Joshi. Covid-19 image classification using deep learning: Advances, challenges and opportunities. *Computers in Biology and Medicine*, page 105350, 2022.

[3] Mohammad A Alzubaidi, Mwaffaq Otoom, and Hamza Jaradat. Comprehensive and comparative global and local feature extraction framework for lung cancer detection using ct scan images. *IEEE Access*, 9:158140–158154, 2021.

[4] S Ashwini, JR Arunkumar, R Thandaiah Prabu, Ngangbam Herojit Singh, and Ngangbam Phalguni Singh. Diagnosis and multi-classification of lung diseases in cxr images using optimized deep convolutional neural network. *Soft Computing*, 28(7):6219–6233, 2024.

[5] Abhir Bhandary, G Ananth Prabhu, Venkatesan Rajinikanth, K Palani Thanaraj, Suresh Chandra Satapathy, David E Robbins, Charles Shasky, Yu-Dong Zhang, João Manuel RS Tavares, and N Sri Madhava Raja. Deep-learning framework to detect lung abnormality–a study with chest x-ray and lung ct scan images. *Pattern Recognition Letters*, 129:271–278, 2020.

[6] Shital D Bhatt and Himanshu B Soni. Image retrieval using bag-of-features for lung cancer classification. In *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, pages 531–536. IEEE, 2021.

[7] Rajendra D Bhosale and DM Yadav. Customized convolutional neural network for pulmonary multi-disease classification using chest x-ray images. *Multimedia Tools and Applications*, 83(6):18537–18571, 2024.

[8] Yogesh H Bhosale and K Sridhar Patnaik. Puldi-covid: Chronic obstructive pulmonary (lung) diseases with covid-19 classification using ensemble deep convolutional neural network from chest x-ray images to minimize severity and mortality rates. *Biomedical Signal Processing and Control*, 81:104445, 2023.

[9] Zenghui Cheng, Yong Lu, Qiqi Cao, Le Qin, Zilai Pan, Fuhua Yan, and Wenjie Yang. Clinical features and chest ct manifestations of coronavirus disease 2019 (covid-19) in a single-center study in shanghai, china. *American Journal of Roentgenology*, 215(1):121–126, 2020.

[10] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[11] Muhammad EH Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al Emadi, et al. Can ai help in screening viral and covid-19 pneumonia? *IEEE Access*, 8:132665–132676, 2020.

[12] Michael Chung, Adam Bernheim, Xueyan Mei, Ning Zhang, Mingqian Huang, Xianjun Zeng, Jiufa Cui, Wenjian Xu, Yang Yang, Zahi A Fayad, et al. Ct imaging features of 2019 novel coronavirus (2019-ncov). *Radiology*, 295(1):202–207, 2020.

[13] Upasana Chutia, Anand Shanker Tewari, Jyoti Prakash Singh, and Vikash Kumar Raj. Classification of lung diseases using an attention-based modified densenet model. *Journal of Imaging Informatics in Medicine*, pages 1–17, 2024.

[14] Abobaker Mohammed Qasem Farhan and Shangming Yang. Automatic lung disease classification from the chest x-ray images using hybrid deep learning algorithm. *Multimedia Tools and applications*, page 1, 2023.

[15] Loveleen Gaur, Ujwal Bhatia, NZ Jhanjhi, Ghulam Muhammad, and Mehedi Masud. Medical image-based detection of covid-19 using deep convolution neural networks. *Multimedia systems*, pages 1–10, 2021.

[16] Zunera Gilani, Yuenting D Kwong, Orin S Levine, Maria Deloria-Knoll, J Anthony G Scott, Katherine L O'Brien, and Daniel R Feikin. A literature review and survey of childhood pneumonia etiology studies: 2000–2010. *Clinical infectious diseases*, 54(suppl_2):S102–S108, 2012.

[17] Abhishek Hazra. A comprehensive survey on chest diseases analysis: technique, challenges and future research directions. *International Journal of Multimedia Information Retrieval*, 10(2):83–110, 2021.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[19] Essam H Houssein, Zainab Abohashima, Mohamed Elhoseny, and Waleed M Mohamed. Hybrid quantum-classical convolutional neural network model for covid-19 prediction using chest x-ray images. *Journal of Computational Design and Engineering*, 9(2):343–363, 2022.

[20] Varadharajan Indumathi and Rathinavelayutham Siva. An efficient lung disease classification from x-ray images using hybrid mask-rcnn and bidlstm. *Biomedical Signal Processing and Control*, 81:104340, 2023.

[21] Sahebgoud Hanamantray Karaddi and Lakhan Dev Sharma. Automated multi-class classification of lung diseases from cxr-images using pre-trained convolutional neural networks. *Expert Systems with Applications*, 211:118650, 2023.

[22] S Kiazyk and TB Ball. Tuberculosis (tb): Latent tuberculosis infection: An overview. *Canada Communicable Disease Report*, 43(3-4):62, 2017.

[23] Stefanus Tao Hwa Kieu, Abdullah Bade, Mohd Hanafi Ahmad Hijazi, and Hoshang Kolivand. A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions. *Journal of imaging*, 6(12):131, 2020.

[24] ZhiFei Lai and HuiFang Deng. Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron. *Computational intelligence and neuroscience*, 2018, 2018.

[25] Suren Makaju, PWC Prasad, Abeer Alsadoon, AK Singh, and A Elchouemi. Lung cancer detection using ct scan images. *Procedia Computer Science*, 125:107–114, 2018.

[26] Darcy D Marciniuk and Dean E Schraufnagel. *The global impact of respiratory disease*. European Respiratory Society, 2017.

[27] Eka Miranda, Mediana Aryuni, and E. Irwansyah. A survey of medical image classification techniques. In *2016 International Conference on Information Management and Technology (ICIMTech)*, pages 56–61, 2016.

[28] Himadri Mukherjee, Subhankar Ghosh, Ankita Dhar, Sk Md Obaidullah, KC Santosh, and Kaushik Roy. Deep neural network to detect covid-19: one architecture for both ct scans and chest x-rays. *Applied Intelligence*, 51(5):2777–2789, 2021.

[29] Md Nahiduzzaman, Md Omaer Faruq Goni, Rakibul Hassan, Md Robiul Islam, Md Khalid Syfullah, Saleh Mohammed Shahriar, Md Shamim Anower, Mominul Ahsan, Julfikar Haider, and Marcin Kowalski. Parallel cnn-elm: A multiclass classification of chest x-ray images to identify seventeen lung diseases including covid-19. *Expert Systems with Applications*, 229:120528, 2023.

[30] Hermann Y Nkouanga and Szilárd Vajda. Automatic tuberculosis detection using chest x-ray analysis with position enhanced structural information. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6439–6446. IEEE, 2021.

[31] World Health Organisation. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023, accessed on 8th may, 2024. *World Health Organisation*, 2023.

[32] Mohd Fauzi Bin Othman, Norarmalina Abdullah, and Nur Aizudin Bin Ahmad Rusli. An overview of mri brain classification using fpga implementation. In *2010 IEEE Symposium on Industrial Electronics and Applications (ISIEA)*, pages 623–628. IEEE, 2010.

[33] Kanika Patra, Sovona Batabyal, Kashmira Mandal, Dhriti Ghose, and Joy Sarkar. Tuberculosis and covid-19: A combined global threat to human civilization. *Clinical Epidemiology and Global Health*, page 101031, 2022.

[34] Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir, Serkan Kiranyaz, Saad Bin Abul Kashem, Mohammad Tariqul Islam, Somaya Al Maadeed, Susu M Zughaier, Muhammad Salman Khan, et al. Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. *Computers in biology and medicine*, 132:104319, 2021.

[35] GV Eswara Rao, B Rajitha, Parvathaneni Naga Srinivasu, Muhammad Fazal Ijaz, and Marcin Woźniak. Hybrid framework for respiratory lung diseases detection based on classical cnn and quantum classifiers from chest x-rays. *Biomedical Signal Processing and Control*, 88:105567, 2024.

[36] Alex Sakula. Robert koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax*, 37(4):246–251, 1982.

[37] FM Javed Mehedi Shamrat, Sami Azam, Asif Karim, Kawsar Ahmed, Francis M Bui, and Friso De Boer. High-precision multiclass classification of lung disease through customized mobilenetv2 from chest x-ray images. *Computers in Biology and Medicine*, 155:106646, 2023.

[38] Dylan Slack, Sophie Hilgard, Sameer Singh, and Himabindu Lakkaraju. How much should i trust you? modeling uncertainty of black box explanations. *arXiv preprint arXiv:2008.05030*, pages 6–25, 2020.

[39] Hai Tang and Zhihui Hu. Research on medical image classification based on machine learning. *IEEE Access*, 8:93145–93154, 2020.

[40] Christian Tchito Tchapga, Thomas Attia Mih, Aurelle Tchagna Kouanou, Theophile Fozin Fonzin, Platini Kuetche Fogang, Brice Anicet Mezatio, and Daniel Tchiotsop. Biomedical image classification in a big data architecture using machine learning algorithms. *Journal of Healthcare Engineering*, 2021, 2021.

[41] Tessa M Wardlaw, Emily White Johansson, and Matthew J Hodge. *Pneumonia: the forgotten killer of children*. Unicef, New York, 2006.

[42] Richard G Wunderink and Grant Waterer. Advances in the causes and management of community acquired pneumonia in adults. *Bmj*, 358, 2017.

[43] Rui Xu, Xiao Cao, Yufeng Wang, Yen-Wei Chen, Xinchen Ye, Lin Lin, Wenchao Zhu, Chao Chen, Fangyi Xu, Yong Zhou, et al. Unsupervised detection of pulmonary opacities for computer-aided diagnosis of covid-19 on ct images. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 9007–9014. IEEE, 2021.

[44] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE transactions on knowledge and data engineering*, 34(12):5586–5609, 2021.

[45] Wentao Zhao, Wei Jiang, and Xinguo Qiu. Deep learning for covid-19 detection based on ct images. *Scientific Reports*, 11(1):14353, 2021.

Research Candidate:

Shazia Mannan,
Research Scholar,
Department of Computer Science,
Mangalore University

Research Supervisor:

Dr. B.H. Shekar,
Professor,
Department of Computer Science,
Mangalore University